DEVELOPMENT STUDIES RESEARCH GROUP

Working Papers

INCOME DISTRIBUTION AMONG SETTLED URBAN AFRICAN HOUSEHOLDS IN SOUTH AFRICA: 1970 AND 1975.

C.E.W. Simkins.

DSRG Working Paper No 6.

University of Natal
Pietermaritzburg
Department of Economics

ACE PMB

Development Studies Research Group

INCOME DISTRIBUTION AMONG SETTLED URBAN AFRICAN HOUSEHOLDS IN SOUTH AFRICA: 1970 AND 1975.

C.E.W. Simkins.

DSRG Working Paper No 6.

Pietermaritzburg
March
1978

TABLE OF CONTENTS

Glossary of Symbols used.

1. Income

x - household income

$z = \ln x$ - natural logarithm of household income, also called income-power

2. Basic statistical concepts

$F(x)$ - value of distribution function at income x

$f(x) = \frac{dF}{dx}$ - value of density function at income x

$F_i = F(x_i)$ - value of distribution function at income class boundary x_i

3. Data

x_i - household income class boundary

k - number of household income classes

n_i - number of households in income class i

μ_i - mean income of households in income class i

N - total number of households in sample

$\bar{x}$ - mean income of all households in sample

4. Cubic spline function

$p_i(x)$ - cubic spline function approximation to $F(x)$ over income class i

$p_i'(x) = \frac{dp_i}{dx}$ - approximation to $f(x)$ over income class i

$h_i = x_{i+1} - x_i$ - width of income class i

$\phi_i = p_i''(x) = \frac{d^2 p_i}{dx^2}$ - second derivative of $p_i(x)$

$\bar{x}_g$ - estimate of the geometric mean of all household incomes

$\tilde{x}$ - estimate of the median household income

5. General inequality measures

s_x - standard deviation of household incomes

$c = \frac{s_x}{\bar{x}}$ - coefficient of variation of household incomes

$i_1 = \frac{1}{1+c^2}$ - normalised coefficient of variation.

s_z - standard deviation of income power

$v = s_z^2$ - variance of income power

$i_2 = \frac{v}{v+1}$ - normalised coefficient of variance of logarithms of income

ε - inequality aversion parameter

i_3 - Atkinson measure with $\varepsilon=1$

i_4 - Atkinson measure with $\varepsilon=2$

i_5 - Gini coefficient

i_6 - normalised Theil measure

6. Analysis of grouping error

J_L - lower limit of general inequality measure J, found by assuming no inequality within income classes

J_U - upper limit of general inequality measure J, found by assuming maximum inequality within income classes

λ_i - proportion of the i th class assumed to receive the lower class income under the maximum inequality assumption.

7. Champernowne parameters

α - parameter mainly determining the distribution of household income among the relatively rich

β - parameter mainly determining the distribution of household income among the middle

γ - parameter mainly determining the distribution of household income among the relatively poor

x_0 - scale parameter

$j_i = \frac{1}{\alpha}$ - inequality measure associated with α i.e. α- inequality or inequality among the relatively rich

$j_2 = \frac{1}{4}(1+\cos(h)_{,}\beta)$ - inequality measure associated with β i.e. β - inequality or inequality among the middle

$j_3 = \frac{1}{\gamma+1}$ - inequality measure associated with γ i.e. γ - inequality or inequality among the relatively poor

s_α, s_γ, standard errors associated with regression estimates of α and γ

$\bar{x}_g^*$ - estimate of geometric mean of household incomes based on Champernowne density function

$\alpha^* = \frac{\gamma+\alpha}{2}$, $\gamma^* = \frac{\gamma-\alpha}{2}$, π_t^*, β_t^*, π_t, β_t, p_t, b_t - auxiliary parameters associated with the Champernowne density function

C - numerator of the Champernowne density function

8. Upper and lower tails

x_h - a high value of household income above which the Pareto law is assumed to hold

$\bar{x}_+$ - mean of all household incomes above x_h

x_l - a low value of household income below which the pseudo-Pareto law is assumed to hold

$\bar{x}_-$ - mean of all household incomes below x_l

9. Goodness-of-fit

χ^2 - value of chi-squared statistic from a goodness of fit test

χ^2_c - critical value of chi-squared at the 5% level of significance

10. Poverty analysis

γ' - estimate of γ assuming that the pseudo-Pareto tail extends to the Minimum Living Level

γ'' - estimate of γ assuming that the pseudo-Pareto tail extends to the median

w_M - proportion of households of size M in total sample of households

y or p - the Minimum Living Level

H - headcount measure of poverty (all households)

H_M - headcount measure of poverty among households of size M

$\bar{x}_y$ - average income of households below the Minimum Living Level

$I = \frac{y-\bar{x}_y}{y}$ - income-gap measure of poverty (all households)

I_M - income-gap measure for households of size M

G - Gini coefficient associated with the distribution of household income among the poor (all households)

G_M - Gini coefficient for poor households of size M

P - Sen poverty measure (all households)

P_M - Sen poverty measure for households of size M

$\frac{\Delta w}{w}$ - proportional wage increase

$\frac{\Delta u}{u}$, $\frac{\Delta v}{v}$ - proportional unemployment increases

b_w, b_u, b_v - sensitivity of H (or P) to $\frac{\Delta w}{w}$, $\frac{\Delta u}{u}$, $\frac{\Delta v}{v}$ respectively

INTRODUCTION.[1]

Income distribution in South Africa may be discussed in three different ways, which focus on, respectively:

(a) the functional distribution of national income between wages and profits

(b) the distribution of personal income between the four racial groups - Whites, Coloureds, Asians and Africans. Often the latter three groups are considered together and then the question becomes that of the distribution of personal incomes between Whites and Blacks. Recent studies suggest that the Black-White distribution was more or less constant over the period from 1925 to 1970, but that it changed between 1970 and 1975 in favour of Blacks.[2] Given that average incomes per capita are very much higher for Whites than for Blacks, this suggests that, in broad terms, the relative position of the poor has improved recently.

(c) the size distribution of personal income either among earners or among households and either within the population as a whole or within particular sections of it.

This study belongs to the third class. It is based on tabulations of income of 'multiple' African households (i.e. African households containing two or more members each) contained in two sets of reports published by the Bureau of Market Research at the University of South Africa.[3] These reports deal with surveys done in 1970 and 1975 in the following cities:

1970	1975
	Bloemfontein
Cape Town	Cape Town
Durban	Durban
East London	East London
Johannesburg	Johannesburg
	Pietermaritzburg
Port Elizabeth/Uitenhage	Port Elizabeth/Uitenhage
Pretoria	Pretoria

The purpose of this study is to assess the degree and type of inequality within the class of settled [4] African households in each of these urban areas and at either date.

The study of size distribution of personal income is important for two reasons: (a) it may yield indicators of the extent of stratification within the population under consideration. Class position is not a function of income alone, but it would be odd to suppose that the class configuration of a population had nothing to do with the size distribution of income.
(b) it serves as a basis for assessing the degree of material poverty within a population using either an absolute level of income as a poverty line (as in this study) or a relative definition of poverty.

A problem arises in the measurement of inequality; recent literature on income distribution [5] has demonstrated that there is no one 'ideal' index. Indices are differentially sensitive to income transfers; some respond most sensitively to changes among the relatively rich, while others register changes among the relatively poor or the middle of the distribution. Accordingly, several measures will be estimated (some in more than one way) and distribution curves will be fitted. A major part of this study (which is a fragment of a much larger project) is the testing of the consistency of various approaches to the study of income distribution - an exercise of a statistical nature.

Interpretation is subordinated to methodological investigation throughout; nonetheless, the results obtained will be interpreted as far as the scope of the study permits. Attempts will be made to discover whether or not there is any systematic variation in inequality over the 1975 cross-section or whether time trends exist. Particular attention will be given to the 'poverty tail' in each case (this is taken to be the part of the distribution in which households have incomes lower than the Minimum Living Level[6]).

There are two reasons in particular, why the interpretation of results should be approached with caution:
(a) Except in the case of one of the approaches to poverty measurement, no attempt is made to move from the distribution of total household incomes to household incomes per consumer unit, i.e. to adjust household incomes for size. The data which would enable one to take this step are not available in published form. Clearly, such an adjustment is desirable; if an inequality measure changes when based on the distribution of total household incomes but remains constant when based on household incomes per consumer unit, the change reflects changing <u>demographic</u> conditions only.[7]

(b) No attempt is made in this study at analysing the causes of changes in inequality. Evaluation of changing inequality depends on the relative weights of factors causing the change; for example, increasing inequality arising solely from increasing discrimination on the grounds of racial identity would generally be regarded as less defensible than increasing inequality based on increased skill differentiation among wage earners.

THEORETICAL DISCUSSION:

Basic statistical concepts and data format

Much of the discussion will involve reference to either:

(i) the distribution function, denoted by $F(x)$ and defined as a probability that a household, drawn at random, will have an income less than or equal to x, or

(ii) the probability density function $f(x) = \frac{dF}{dx}$

It will be assumed throughout that negative incomes are impossible;[8] that $F(x)$ increases monotonically over the interval $[0,\infty)$ with $F(0)=0$ and $F(\infty)=1$ and that $F(x)$ is continuous and differentiable over this interval. Then $f(x)$ exists and is non-negative over $[0,\infty)$
The data comes in the form : n_i households have incomes in the range

$[x_1 = 0, x_2)$; n_2 in $[x_2,x_3),\ldots,$ n_k in $[x_k,\infty)$.

If $N = \sum_{i=1}^{k} n_i$ the total number of households, then $\frac{\sum_{i=1}^{r-1} n_i}{N}$ the cumulated proportion of households in the first r-1 income ranges, can be taken as an observation of the distribution function at x_r i.e. $F(x_r)$.

At each intermediate base point $x_2,\ldots,x_k$, 95% confidence limits for $F_i=F(x_i)$ can be calculated since F_i can be regarded as an estimate of the proportion of households with incomes at or below x_i. The standard error of this estimate given a sample size N is $\sqrt{\frac{F_i(1-F_i)}{N}}$ and the confidence interval becomes

$$\left(F_i-1{,}96\sqrt{\frac{F_i(1-F_i)}{N}},\quad F_i+1{,}96\sqrt{\frac{F_i(1-F_i)}{N}}\right)$$ [9]

The cubic spline function and general inequality measures.

We start by finding the approximation to $F(x)$ for other values of x. Over the range $[0, x_k]$ this involves interpolation; extrapolation is required over the range (x_k, ∞). The technique used is that of construction of a cubic spline function [10] which is then modified. Details are discussed in Appendix I.

An arithmetic mean may be calculated from the spline function.

Since $\bar{x} = \int_{x_1}^{x_n} xf(x)dx$, $\bar{x} = \sum_{i=1}^{k} \int_{x_i}^{x_{i+1}} x\, p_i'(x)dx$. Similarly the

geometric mean $\bar{x}_g$ is given by $\bar{x}_g = \exp\left[\sum_{i=1}^{k} \int_{x_i}^{x_{i+1}} \ln x\, p_i'(x)dx\right]$

Five normalised (in the sense that they vary from zero when all households have equal incomes to one when one household has all the income) inequality measures are also computed.

These are:

- the normalised coefficient of variation (i_1)
- the normalised coefficient of variance of logarithms (i_2)
- the Atkinson inequality measure with inequality aversion parameter $\varepsilon=1$ (i_3)
- the Atkinson measure with $\varepsilon=2$(i_4)
- the normalised Theil measure (i_6).[11]

All these are relative measures in the sense that they would remain constant if all incomes were, say, doubled or halved. The first two are based on the ratios between the standard deviations and the means of income and of the logarithm of income (sometimes known as income-power) respectively. The Atkinson measure is based explicitly on social welfare function considerations; the concept of the equally distributed equivalent level of income (y_{EDE}) is introduced. This is the level of income per household which if equally distributed would give the same level of social welfare as the distribution under consideration. The inequality measure is defined as one minus the ratio of y_{EDE} to the mean of the actual distribution. An interpretation of the measure then suggests itself; if it is 0,2, for instance, only 80% of total income would be needed to achieve the same level of social welfare provided that income were equally distributed. Now, if one assumes that the social welfare function is symmetric (i.e. an income of x makes the same contribution to social welfare no matter to which household it accrues) and additively separable (i.e. the utility of a household depends on its income alone and social welfare is the sum of household utilities) and if one requires the inequality measure to be a relative one, it turns out that the form of the social welfare measure and hence, the inequality measure, is defined up to a non-negative parameter ε.[12] ε, called the inequality-aversion parameter by Atkinson, has a relatively simple interpretation - the marginal utility of income is inversely proportional to income raised to the power ε. If $\varepsilon=1$, the marginal utility of income is inversely proportional to the level of income; if $\varepsilon=2$ it is inversely proportional to the square of income.

The Theil measure is based on information-theoretic considerations. It was initially based on the entropy of the income shares of N individuals (households in this case); complete equality would give an entropy of ln N while complete inequality would give an entropy of zero. Theil converts this into an inequality measure by subtracting entropy from its maximum value. This measure is already a relative one, being based on income shares, but has to be normalised.

The variance of logarithms and the un-normalised Theil measure have the advantage that they can be decomposed into between-set and within-set variations. This property is not exploited in the present study, but would be useful in a discussion of the various components of inequality.

Formulae used in the computation of the various general inequality measures are discussed in Appendix II.

Grouping and sampling errors

If we are supplied with class means from the data source, it is possible to compute the upper and lower limits consistent with the data without curve fitting for each of these measures. Denote the measure by J, the lower limit by J_L and the upper limit by J_U. J_L is found by assuming that everyone in each class gets the mean income in that class i.e. J_L is computed on the assumption that there is no inequality within the classes. J_U is found by assuming that there is maximum inequality within each class; this situation occurs when households are assumed to receive incomes at the lower and upper class limits in such proportions as are necessary to reproduce the class mean. The range between J_L and J_U is called the grouping error.

If λ_i denotes the proportion of the i th class assumed to receive the lower limit income, $\lambda_i = \frac{x_{i+1}-\mu_i}{x_{i+1}-x_i}$ where μ_i is the class mean.[13] If the class means are not given, they can be calculated from $\mu_i = \frac{\int_{x_i}^{x_{i+1}} x p'_i(x)dx}{F(x_{i+1})-F(x_i)}$

If J_L and J_U are calculated from empirically given class means and J from the cubic spline function, J may lie slightly outside J_L, J_U since the spline function will not produce the empirical means exactly. Upper and lower bounds on the Gini coefficient, i_5, may be calculated in this way. We do not obtain a point estimate of i_5 from the cubic spline function since this involves a double numerical integration; instead i_5 is estimated as $\frac{2}{3}(i_5)_U + \frac{1}{3}(i_5)_L$, found by experience to be a good estimator. The Gini coefficient is, of course, the best known measure of inequality, being the ratio of the area between the line of perfect equality and the Lorentz curve and the total area under the line of perfect equality in a Lorentz diagram. It has the demerit of always being more difficult to estimate than the other measures and is included only because of its familiarity.

Since we are working with sample surveys (and the sample size is relatively small), we should also consider the standard errors of our estimates (sampling error) where possible. This can be done for i_1 via the associated variable c, for the coefficient of variation for i_2 via the variance of logarithms S_z^2 and for i_5. The standard error in c is approximately $c\sqrt{\frac{1+2c^2}{N}}$, in $S_z^2 = v$ approximately $v\sqrt{\frac{2}{N}}$ and in i_5 approximately $i_5\sqrt{\frac{1+2c^2}{2N}}$[14]. Approximate 95% confidence limits are found by computing bounds two standard errors on either side of the estimate.

The Champernowne density function.

Instead of fitting a cubic spline distribution function to the data, we may attempt to fit one of the 'standard' income distribution curves. Champernowne [15] suggests a probability density function of the form

$$f(x) = \frac{C}{x\left[\left(\frac{x}{x_0}\right)^{-\gamma} + 2\cos\beta \left(\frac{x}{x_0}\right)^{-\gamma^*} + \left(\frac{x}{x_0}\right)^{\alpha}\right]} \qquad x > 0$$

$$\text{where } \gamma^* = \frac{\gamma \ \alpha}{2} \quad \text{and} \quad C = \frac{\alpha^* \sin\beta \sin\left(\frac{\gamma^*\pi}{\alpha^*}\right)}{\pi \sin\left(\frac{\gamma^*\beta}{\alpha^*}\right)} \quad \text{and } \alpha^* = \frac{\gamma+\alpha}{2}$$

which depends on three parameters α, β, γ, and a fourth scale parameter x_0. The interest of the Champernowne density function lies in the fact that α, β and γ may be associated with different parts of the distribution. Consider the portion of the denominator within square brackets. When x is small compared with the scale parameter x_0, the term $\left(\frac{x}{x_0}\right)^{-\gamma}$ dominates the other two, so γ is associated with the bottom end of the distribution. When x is large compared with x_0, the term $\left(\frac{x}{x_0}\right)^{\alpha}$ dominates, so α is associated with the top end of the distribution. β is associated with middle incomes. Accordingly, we may define three 'specialist' inequality measures $j_1 = \frac{1}{\alpha}$, $j_2 = \frac{1}{\pi}(1 + \cos\beta)$ and $j_3 = \frac{1}{\gamma+1}$ which measure inequality amongst the rich, the middle income receivers and poor, respectively. As before, the higher these measures, the greater the inequality. Note that j_1 (which may also be called α- inequality) and j_3 (γ- inequality) are inversely related to α and γ respectively. The lower are α and γ, the greater α- and γ- inequality.

The problem to be tackled here is that of estimating the parameters from the data. We proceed as follows:

(i) When x is large $f(x) \simeq \dfrac{C}{x\left[\dfrac{x}{x_0}\right]^{\alpha}} = \dfrac{Cx_0^{\alpha}}{x^{\alpha+1}}$

This is the Pareto density function; we know that the proportion of incomes exceeding x is proportional to $x^{-\alpha}$ (Pareto's law)

i.e. $1-F(x) = kx^{-\alpha}$ for large x

$\ln(1-F(x)) = \ln k - \alpha \ln x$

and so α may be estimated using ordinary least squares regression. To ensure x is as large as possible, we confine our attention to the three points $(x_{n-3}, F(x_{n-3}))$, $(x_{n-2}, F(x_{n-2}))$ and $(x_{n-1}), F(x_{n-1}))$

When x is small $f(x) \simeq \dfrac{C}{x\left[\dfrac{x}{x_0}\right]^{-\gamma}} = \dfrac{C}{x_0^{\gamma}} x^{\gamma-1}$

and $F(x) = \int_0^x \dfrac{C}{x_0^{\gamma}} v^{\gamma-1} dv = \dfrac{C}{x_0^{\gamma}} \dfrac{x^{\gamma}}{\gamma}$,

so F(x) is proportional to x^{γ} (this may be called the pseudo-Pareto law)

i.e. $F(x) = k^*x^{\gamma}$

$\ln F(x) = \ln k^* + \gamma \ln x$

and γ may also be estimated using OLS regression. To ensure x is as small as possible, the three points $(x_2, F(x_2))$, $(x_3, F(x_3))$ and $(x_4, F(x_4))$ only are considered.

(ii) This leaves β and x_0 to be estimated. The arithmetic mean (either supplied or computed from the spline function) and the median (interpolated from the spline function) are used for this purpose.
Champernowne[16] defines the auxiliary parameters

$\pi_t^* = \dfrac{(\gamma^*+t)\,\pi}{\alpha^*}$ $\qquad \beta_t^* = \dfrac{(\gamma^*+t)\beta}{\alpha^*}$

$\pi_t = \pi_t^* \cot \pi_t^*$ $\qquad \beta_t = \beta_t^* \cot \beta_t^*$ $\qquad \beta_t = 1$ if $\beta_t^* = 0$

$p_t = \pi_t^* \operatorname{cosec} \pi_t^*$ $\qquad b_t = \beta_t^* \operatorname{cosec} \beta_t^*$ $\qquad b_t = 1$ if $\beta_t^* = 0$

and then shows[17] that $\dfrac{\bar{x}}{x_0} = \dfrac{b_0 p_1}{b_1 p_0}$

Denote the median by $\tilde{x}$, then $\int_0^{\tilde{x}} \frac{C\,dx}{x\left[\left(\frac{x}{x_o}\right)^{-\gamma} + 2\cos\beta\left(\frac{x}{x_o}\right)^{-\gamma^*} + \left(\frac{x}{x_o}\right)^{\alpha}\right]} = \frac{1}{2}$

Now,

$$\frac{x}{x_o} = \frac{x}{\bar{x}}\,\frac{\bar{x}}{x_o} = \frac{x}{\bar{x}}\,\frac{b_o p_1}{b_1 p_o}, \text{ so}$$

$$\int_0^{\tilde{x}} \frac{C\,dx}{x\left[\left(\frac{x}{\bar{x}}\,\frac{b_o p_1}{b_1 p_o}\right)^{-\gamma} + 2\cos\beta\left(\frac{x}{\bar{x}}\,\frac{b_o p_1}{b_1 p_o}\right)^{-\gamma^*} + \left(\frac{x}{\bar{x}}\,\frac{b_o p_1}{b_1 p_o}\right)^{\alpha}\right]} - \frac{1}{2} = 0$$

This equation is solved numericallyfor β by the half-interval method, since all other variables in the integrand ($\bar{x}$, α, γ)are known and $\tilde{x}$ is also known:[18] Once β is found, we substitute back into the equation for the mean to find x_o.

(iii) Curves estimated in this fashion were found not to fit the data satisfactorily, so an additional refinement was introduced. Champernowne[19] shows that the geometric mean $\bar{x}_g$ is given by

$$\frac{\bar{x}_g}{x_o} = \exp\left\{\frac{\beta_o - \pi_o}{\gamma^*}\right\}$$

We have an estimate of $\bar{x}_g$ from the spline function work, so we adjust α and γ in such a way as to reproduce it. Since α and γ are estimated by regression, the standard errors of α and γ can be estimated and the adjustments are made in such a way that they are proportional to the standard error in each case. Adjusting α and γ, of course, involves adjusting β and x_o as well.[20] In some cases a yet better fit can be found by adjusting γ only.

Alternative estimates for α and γ may be found by using the formulae $1 - j_1 = 1 - \frac{1}{\alpha} = \frac{x_h}{\bar{x}_+}$ where x_h is any high income and $\bar{x}_+$ is the arithmetic mean of incomes exceeding x_h i.e. $\alpha = \frac{x_h}{\bar{x}_+ - x_h}$ and $1 - j_3 = 1 - \frac{1}{\gamma+1} = \frac{\bar{x}_-}{x_1}$ where x_1 is any low income and $\bar{x}_-$ is the arithmetic mean of incomes less than x_1.[21] i.e. $\gamma = \frac{\bar{x}_-}{x_1 - \bar{x}_-}$

The general inequality measures i_1, i_2, i_3, i_4 and i_6 can be calculated from the auxiliary parameters (i.e. ultimately from α, β and γ)[22]. The three specialist inequality measures j_1, j_2, j_3 are readily calculated. Finally, numerical integration may be used to find the expected proportion (and hence number) of households in each class; a χ^2 statistic may be constructed as a measure of the goodness of fit of the Champernowne curve to the data.

Comparison of estimates

Graphical illustration of the goodness of fit is supplied by plotting on one graph

(i) the empirically given points on the distribution function
(ii) 95% confidence limits on either side of these points, calculated as described above
(iii) the Champernowne distribution function

The following chart summarises the point and interval estimates obtained for the various inequality measures :

<table>
<tr><th>Measure</th><th>cubic spline distribution function - point estimate</th><th>Grouping limits - upper and lower</th><th>95% confidence limits - sampling error</th><th colspan="2">Champernowne density function - point estimate</th></tr>
<tr><td>i_1</td><td>✓</td><td>✓</td><td>✓</td><td colspan="2">✓</td></tr>
<tr><td>i_2</td><td>✓</td><td>✓</td><td>✓</td><td colspan="2">✓</td></tr>
<tr><td>i_3</td><td>✓</td><td>✓</td><td></td><td colspan="2">✓</td></tr>
<tr><td>i_4</td><td>✓</td><td>✓</td><td></td><td colspan="2">✓</td></tr>
<tr><td>i_5</td><td></td><td>✓</td><td>✓</td><td colspan="2"></td></tr>
<tr><td>i_6</td><td>✓</td><td>✓</td><td></td><td colspan="2">✓</td></tr>
<tr><td></td><td></td><td></td><td></td><td>regression</td><td>alternative method</td></tr>
<tr><td>j_1</td><td></td><td></td><td></td><td>✓</td><td>✓</td></tr>
<tr><td>j_2</td><td></td><td></td><td></td><td colspan="2">✓*</td></tr>
<tr><td>j_3</td><td></td><td></td><td></td><td>✓</td><td>✓</td></tr>
</table>

*Note: j_2 is, of course, obtained neither by regression nor by the alternative method.

An important test of the consistency of our estimation procedures will be to compare estimates of the general inequality measures where these have been obtained by different procedures.

Poverty analysis

It remains to discuss the poverty analysis. Given a poverty datum line at income y, one may construct four measures:

(i) H, the headcount measure, defined as the proportion of total households with incomes less than y. Sen [23] criticises this measure on two counts:

(a) an unchanged number of people below the poverty line may go with a sharp change in the extent of the shortfall in income from the line.

(b) it is completely insensitive or responds perversely to the distribution of income among the poor.

The next two measures (separately) are designed to meet these criticisms:

(ii) I, the income-gap measure, defined as the proportion of the poverty level by which the average household lying below this level falls short of it.

(iii) G, the Gini coefficient of the income distribution <u>of the poor</u>

(iv) The final measure considered is Sen's poverty index P. Sen puts [24] $P = H\,[I + (1-I)\,G]$ for large numbers of the poor. He provides interpretation of P in the following terms: " I represents poverty as measured by the proportionate gap between the mean income of the poor and the poverty line income. It ignores the distribution <u>among</u> the poor and G provides this information. In addition to the poverty gap of the mean income of the poor reflected in I, there is the 'gap' arising from the unequal distribution of the mean income, which is reflected by the Gini coefficient G of that distribution multiplied by the mean income ratio. The income-gap measure thus augmented to take note of the inequality among the poor i.e. $I + (1-I)\,G$ is normalised per poor person, and does not take note of the number of people below the poverty line, which could be minute or large. Multiplying $I + (1-I)\,G$ by the headcount ratio H now produces the composite measure P." [25] More fundamentally, Sen proves that P is the only measure having three desirable properties (a detailed discussion of which is not intended here).

This discussion implicitly assumes that there is a single poverty datum line for <u>all</u> households. If, however, the line is constructed by costing a collection of goods and services necessary for subsistence (however defined), then it will vary with household size. In fact, the Minimum Living Level is calculated for households of sizes 2,3,4,5,6,7 and 8+, as well as for a household of average size. Given the form of the data, which does not contain separate income distribution tabulations for households of various sizes, two approaches to the measurement of poverty suggest themselves. Neither is completely satisfactory.

H is obtained by interpolation on the cubic spline distribution function i.e. $H = F(y) \approx p_i(y)$ for the appropriate i. i.e. y in $[x_i, x_{i+1}]$.

In order to estimate G and I the further simplifying assumption is made that the distribution function for households below the MLL is of the form

$$F(x) = Ax^{\gamma'} \quad (0 \leq x \leq y)$$

i.e. that the psuedo-Pareto law holds up to the MLL.

The bigger H, the less accurate will this be as an approximation.
Then it can be shown (see Appendix III) that

$$I = \frac{1}{\gamma' + 1} \qquad G = \frac{1}{2\gamma' + 1}$$

and

$$P = H \frac{3\gamma' + 1}{(\gamma' + 1)(2\gamma' + 1)}$$

Like γ, γ' is estimated by ordinary least squares regression, but the points considered now range from $(x_2, F(x_2))$ to $(x_m, F(x_m))$ where x_m is the base-point nearest to y.

The second approach seeks to generate separate income distributions for households in each size category. The available information which serves as a basis for this procedure is:

- a table of the average number of earners [26] per household for households of different sizes
- a table of the number of households in each size category
- distributions of earnings of men and women.

The generation procedure is as follows:

- households of size M = 2,3, ... 8 are successively split up into groups having 1,2, ... M earners on the assumption that the probability of a household of size N having r earners is represented by a modified binomial probability mass function.[27]
- it is assumed that the proportion of men and women among third and subsequent earners (in households which have them) is the same as the proportion of men and women earners as a whole. Second earners are assumed all to be women. This leaves a residue of both men and women (more of the former) who are the first earners
- for households of size M lists of income are built up successively; households having 1 earner are considered first, then those with 2 and so forth. Earnings are drawn randomly from the distribution of earnings for

men and women in accordance with the postulated proportion of men and women among 1st, 2nd ... etc. earners.

- once a list of incomes for all households of size M is compiled, the rest is straightforward. The proportion falling below the MLL can be established (H_M), and the income-gap measure (I_M) and Gini coefficient among the poor G_M can be calculated and the Sen Poverty measure (P_M) computed. If the proportion of households of size M in the total sample is w_M, then H and P can be calculated from the simple formulae:

$$H = \Sigma\, w_M H_M$$

$$P = \Sigma\, w_M P_M$$

Two tests are available for determining whether this procedure leads to results consistent with all the available information:

(i) a distribution of <u>all</u> households by number of earners is available and this can be compared with the sum of households having 1,2, ... etc. earners in households of sizes 2,3, ... 8+.

(ii) the distribution of income among households of sizes 2,3 ... etc. may be pooled to give a distribution of income for households of all sizes and this can be compared with the tabulated distribution.

Having built up this framework, one may go a step further and assess the effects of (a) higher wages and (b) a lower level of employment on H and P. All earners were given a 1,2, ...8 percent increase in wages and H and P recalculated. <u>Two</u> ways of introducing lower employment were considered:

(i) the average number of earners in each household size category was lowered by 2,4, ... 16 percent,[28] and H and P recalculated. The effect of this is to knock out second and subsequent earners, but never first earners since the modified binomial distribution does not allow for zero earners.

(ii) incomes of 1,2, ... 8 percent of earners (whether first or subsequent) were reduced to zero and H and P recalculated.

Nine points for wage increases and nine each for the two types of employment drop were thus generated. In each case, H and P were regressed on the wage increase/employment drop using a simple linear model e.g.

$$H = H_o + b_w \frac{\Delta w}{w} \qquad \text{(for wage increases)}$$

or

$$\Delta H = H - H_o = b_w \frac{\Delta w}{w}$$

Considering the situation from a static point of view, we would expect employment to drop as a consequence of a wage increase, the proportional size of the drop to the proportional size of the increase being the wage elasticity of demand for labour. Now, ΔH will be positive if $\frac{\Delta w}{w}$ is positive and negative if $\frac{\Delta u}{u}$ (the proportional size of the employment drop) is positive. A certain size of $\frac{\Delta w}{w}$ will therefore balance a certain size of $\frac{\Delta u}{u}$ in the sense that, if both are applied, H will be unchanged. We therefore arrive at the notion of a critical wage elasticity of demand for labour, given by $\left|\frac{b_u}{b_w}\right|$; if the actual elasticity is above this level, then an increase in wages can be expected to worsen poverty because of the associated drop in employment. Critical elasticities may be derived for both H and P and for both types of unemployment incidence.

Computer programmes have been written to carry out most of the calculations discussed above.

RESULTS:

Throughout this section the unit of income measurement, is one thousand rand per year.

Curve fitting and inequality measures.

The means generated from the cubic spline distribution function and the empirically given means are compared in Table I.

TABLE I

COMPARISON OF MEAN HOUSEHOLD INCOMES IN 1975.

City	Cubic spline function mean (kR/year)	Empirical mean (kR/year)	Deviation (%)
Bloemfontein	1,276	1,265	+ 0,87
Cape Town	2,062	2,077	- 0,72
Durban	2,125	2,130	- 0,23
East London	1,615	1,625	- 0,62
Johannesburg	2,258	2,272	- 0,62
Pietermaritzburg	2,328	2,394	- 2,76
Port Elizabeth	1,933	1,938	- 0,26
Pretoria	1,953	1,990	- 1,86
		Mean	- 0,78
		Standard deviation	1,10

Generally, then the agreement between the estimated and supplied means is quite good. A similar result was found for 1970 where the mean deviation was - 0,93% and the standard deviation of the deviations 1,75%.

Values of α, β, γ and x_0 are tabulated for each city in 1975 in Table II along with the corresponding specialist inequality indices j_1, j_2, j_3.

TABLE II.

CHAMPERNOWNE PARAMETERS AND SPECIALIST INEQUALITY MEASURES IN 1975.

City	α	β: $\cos\beta$ (form of density function)	$\cosh\beta$	γ	x_0	$\frac{1}{\alpha}$ j_1 (α-inequality)	$\frac{1}{4}(1+\cos(h)\beta)$ j_2 (β-inequality)	$\frac{1}{\gamma+1}$ j_3 (γ-inequality)
Bloemfontein	2,94	2,05		2,17	1 222	0,340	0,135	0,315
Cape Town	4,17	0,77		2,07	2,493	0,239	0,429	0,325
Durban	3,90	0,83		2,65	2,207	0,256	0,419	0,273
East London	5,46		2,30	2,07	2,535	0,183	1,512	0,325
Johannesburg	3,39	1,68		2,61	2,207	0,295	0,222	0,277
Port Elizabeth	3,54	0,31		2,73	1,875	0,282	0,488	0,267
Pretoria	3,46	0,81		2,23	2,080	0,289	0,421	0,309

Note: It proved to be impossible to fit a curve to the data for Pietermaritzburg.

Values of α range from 2,94 to 5,46, indicating a generally low (and sometimes very low) degree of inequality among the relatively rich. In the United Kingdom and other western European countries, α is around 2,5; in some Eastern European countries this rises to 4,0 - 4,5. The range for γ is smaller (2,07 to 2,73) while β- inequality shows the biggest range of all, ranging from very small in Bloemfontein to very large in East London.

α and γ may be estimated in three different ways (discussed above):

(i) by OLS regression only

(ii) by OLS regression and then adjustment to reproduce the geometric mean

(iii) from means of low incomes and high incomes respectively. The three sets of estimates are compared in Tables III A and III B.

TABLE III A.

COMPARISON OF ESTIMATES OF α in 1975

City	Regression (A)	Adjusted regression (B)	x_h	$\bar{x}_+$	$\alpha = \frac{x_h}{\bar{x}_+ - x_h}$ (C)	City ranks by estimates of α		
						R_A	R_B	R_C
Bloemfontein	3,35	2,94	2,000	3,044	1,92	6	7	7
Cape Town	4,17	4,17	4,000	4,825	4,85	2	2	1
Durban	3,87	3,90	4,000	5,309	3,06	3	3	3
East London	5,46	5,46	4,000	5,556	2,57	1	1	4
Johannesburg	3,39	3,39	3,000	4,212	2,48	5	6	5
Port Elizabeth	3,78	3,54	4,000	4,997	4,01	4	4	2
Pretoria	2,72	3,46	4,000	5,629	2,46	7	5	6
Median	3,78	3,54			2,57			

TABLE III B.

COMPARISON OF ESTIMATES OF γ in 1975

City	Regression (A)	Adjusted regression (B)	x_1	$\bar{x}_-$	$\gamma = \frac{\bar{x}_-}{x_1 - \bar{x}_-}$ (C)	City ranks by estimates of γ		
						R_A	R_B	R_C
Bloemfontein	2,04	2,17	0,750	0,496	1,95	5	5	6
Cape Town	1,41	2,07	1,250	0,854	2,16	7	7	3
Durban	2,90	2,65	1,250	0,841	2,06	1	2	5
East London	2,20	2,07	1,000	0,633	1,72	4	6	7
Johannesburg	1,98	2,61	1,750	1,186	2,10	6	3	4
Port Elizabeth	2,65	2,73	1,000	0,689	2,22	3	1	1
Pretoria	2,83	2,23	1,000	0,686	2,18	2	4	2
Median	2,20	2,23			2,10			

In the case of the α's, the estimates based on the averages of high incomes yield lower values than those based on regression. In the case of the γ's, the estimates are closer together. Another test of the consistency of the estimates is to test how consistently they rank the seven cities; the Spearman rank correlation coefficient is suited to this purpose.

TABLE IV.

SPEARMAN RANK CORRELATION COEFFICIENTS BETWEEN METHODS A, B AND C OF ESTIMATING α AND γ IN 1975.

	S_{AB}	S_{BC}	S_{AC}
α	0,893	0,714	0,714
γ	0,607	0,429	0,107

The 5% significance level (one-tailed test) for S if there are seven observations is 0,714, so we conclude that there is consistency in ranking for the estimates of α but not for γ. These results suggest that there are significant differences in the α's between cities; comparisons of the median raise the possibility that the values of α may be generally somewhat lower than the adjusted regression estimates; the differences in γ are less likely to be significant, in most cases, the estimates yield a value of γ a little above 2.[29]

How well do the Champernowne curves fit the data? On figures 1 to 7 are plotted:

(i) the empirical distribution function, denoted by -x-x-

(ii) 95% confidence limits for the distribution function, denoted by -+-+-

(iii) the Champernowne distribution function -o-o-

The χ^2 goodness-of-fit statistic is compared with the critical value of χ^2 for lack of fit at the 5% level in Table V.

FIGURE 2.

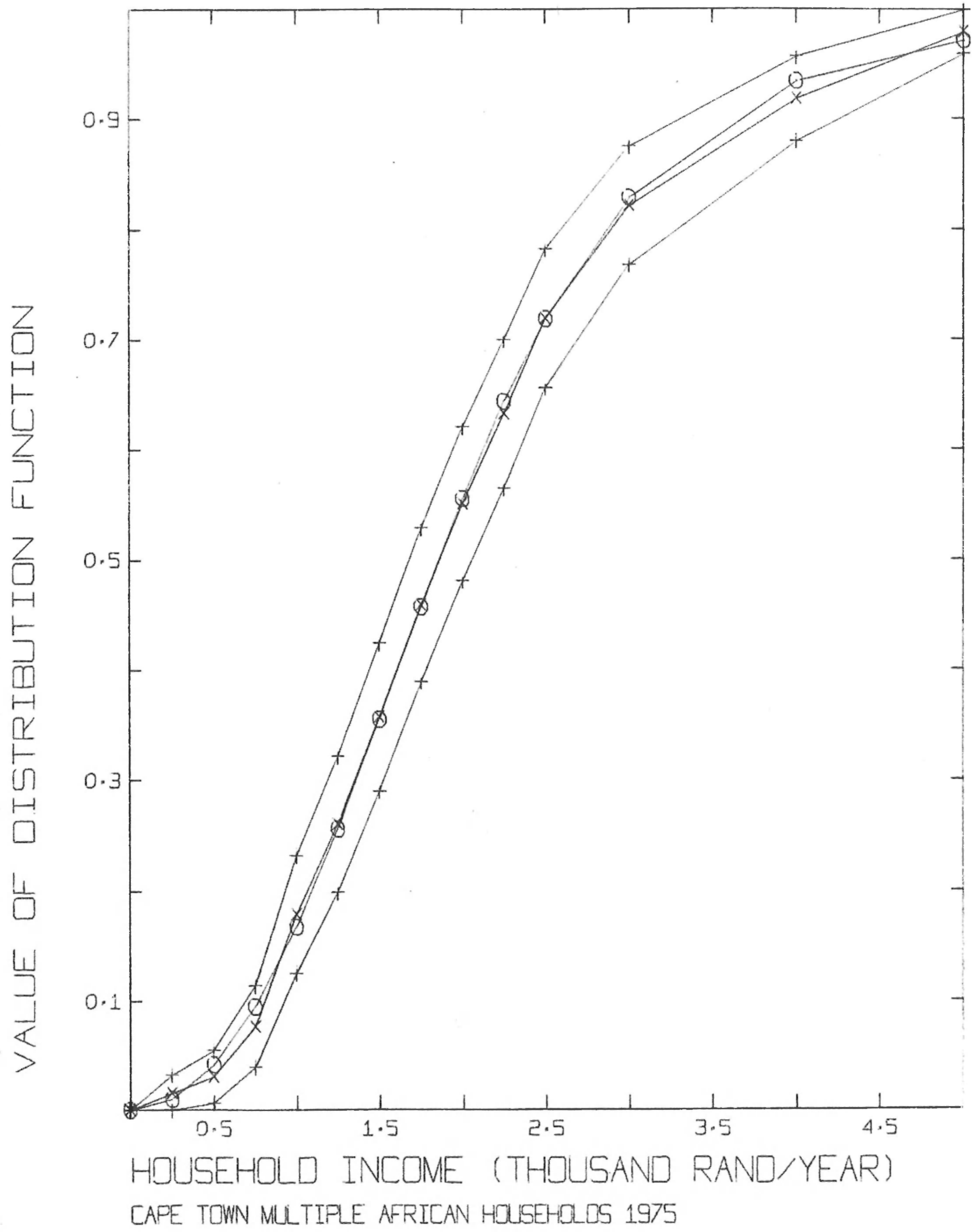
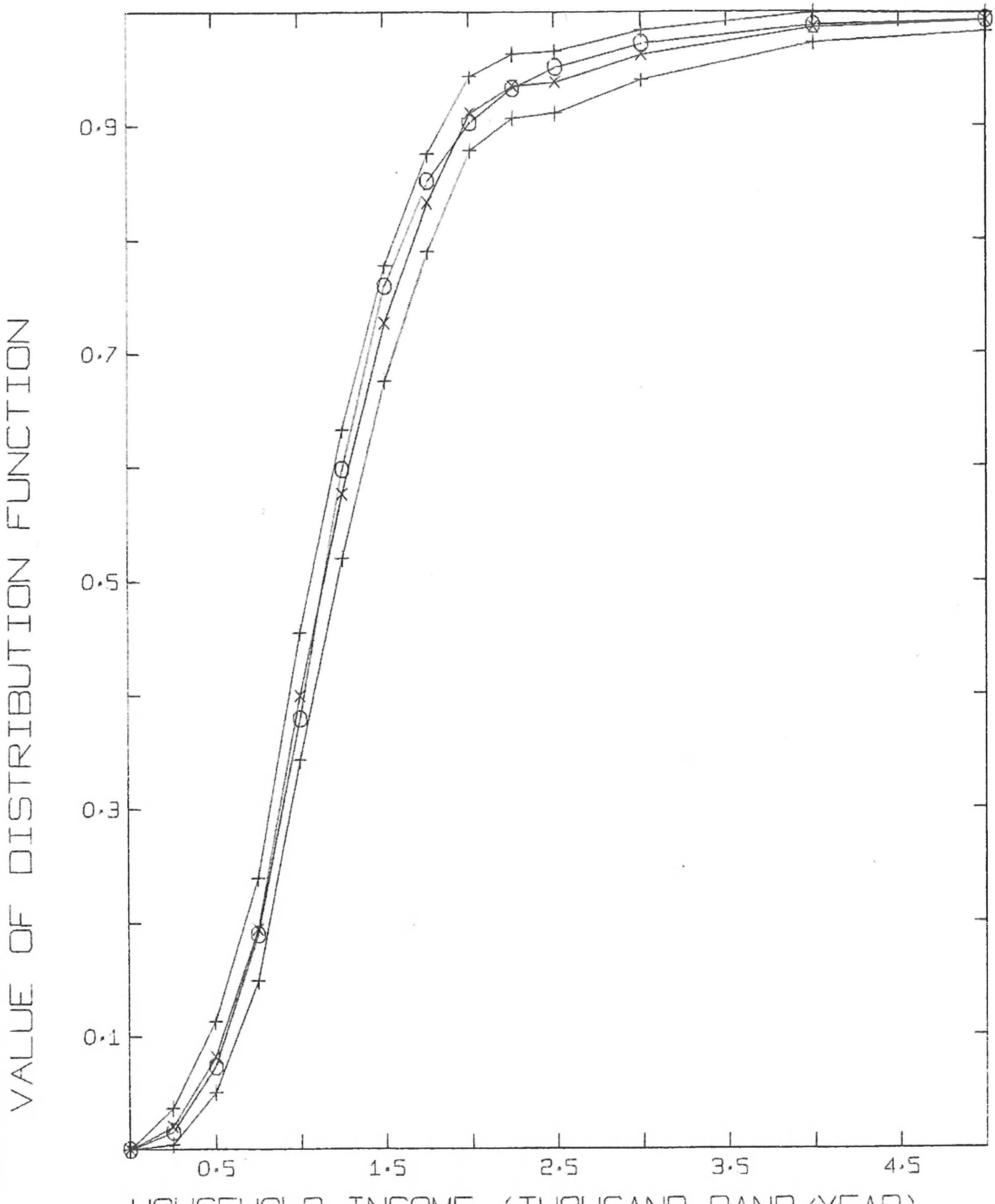
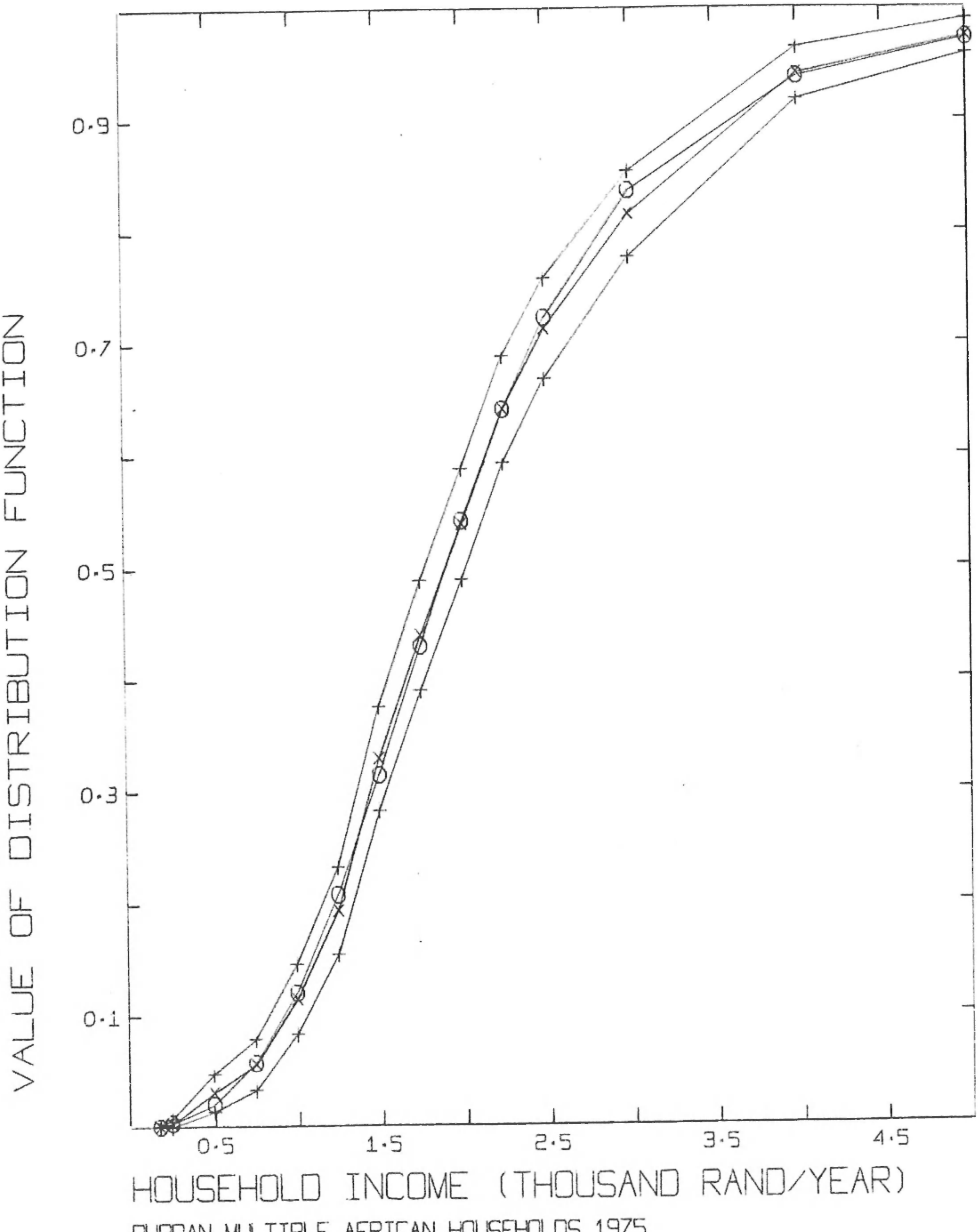

VALUE OF DISTRIBUTION FUNCTION
0·9
0·7
0·5
0·3
0·1
0·5
1·5
2·5
3·5
4·5
HOUSEHOLD INCOME (THOUSAND RAND/YEAR)
CAPE TOWN MULTIPLE AFRICAN HOUSEHOLDS 1975

FIGURE 1.



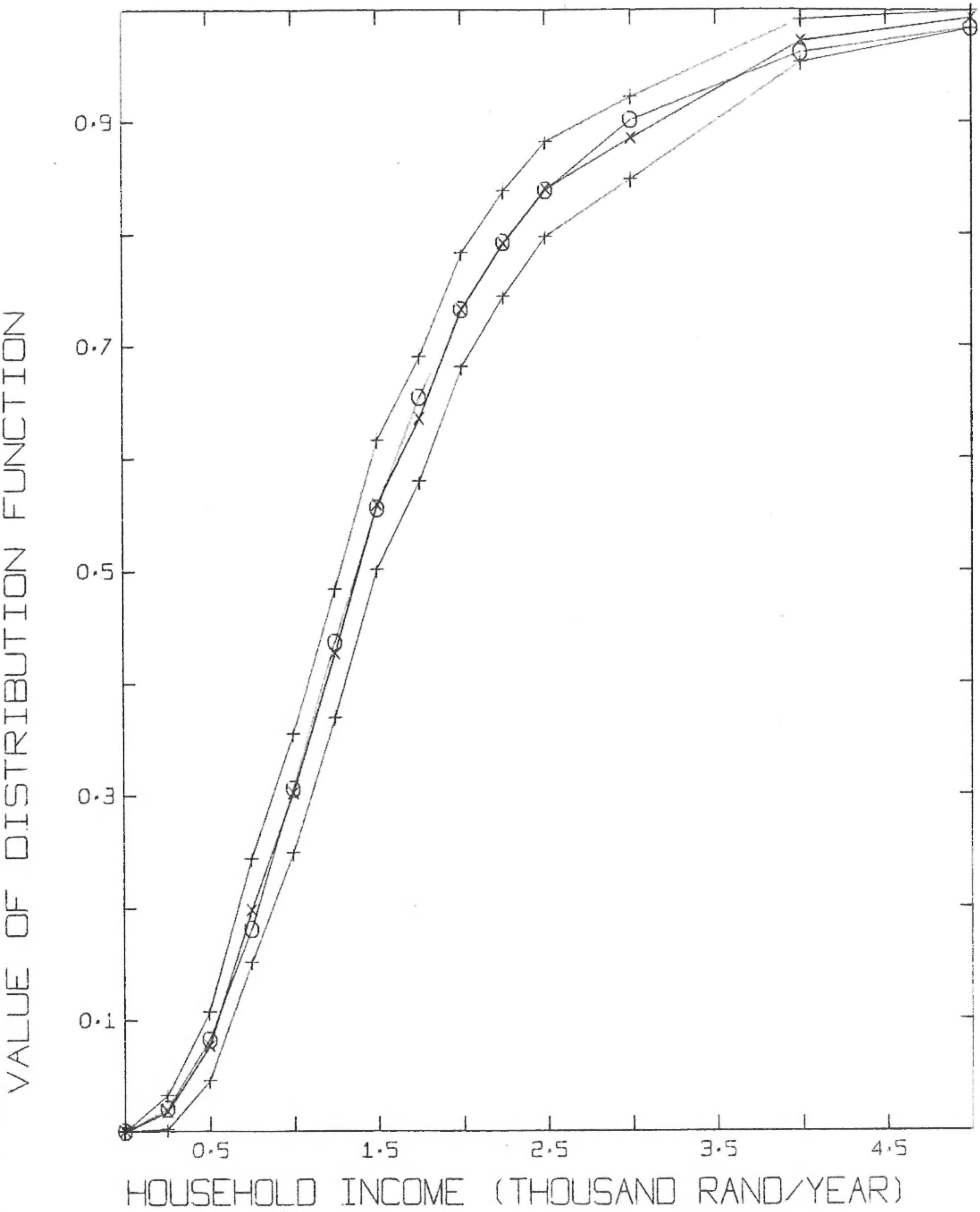
BLOEMFONTEIN MULTIPLE AFRICAN HOUSEHOLDS 1975

FIGURE 3.



DURBAN MULTIPLE AFRICAN HOUSEHOLDS 1975

FIGURE 4.



EAST LONDON MULTIPLE AFRICAN HOUSEHOLDS 1975

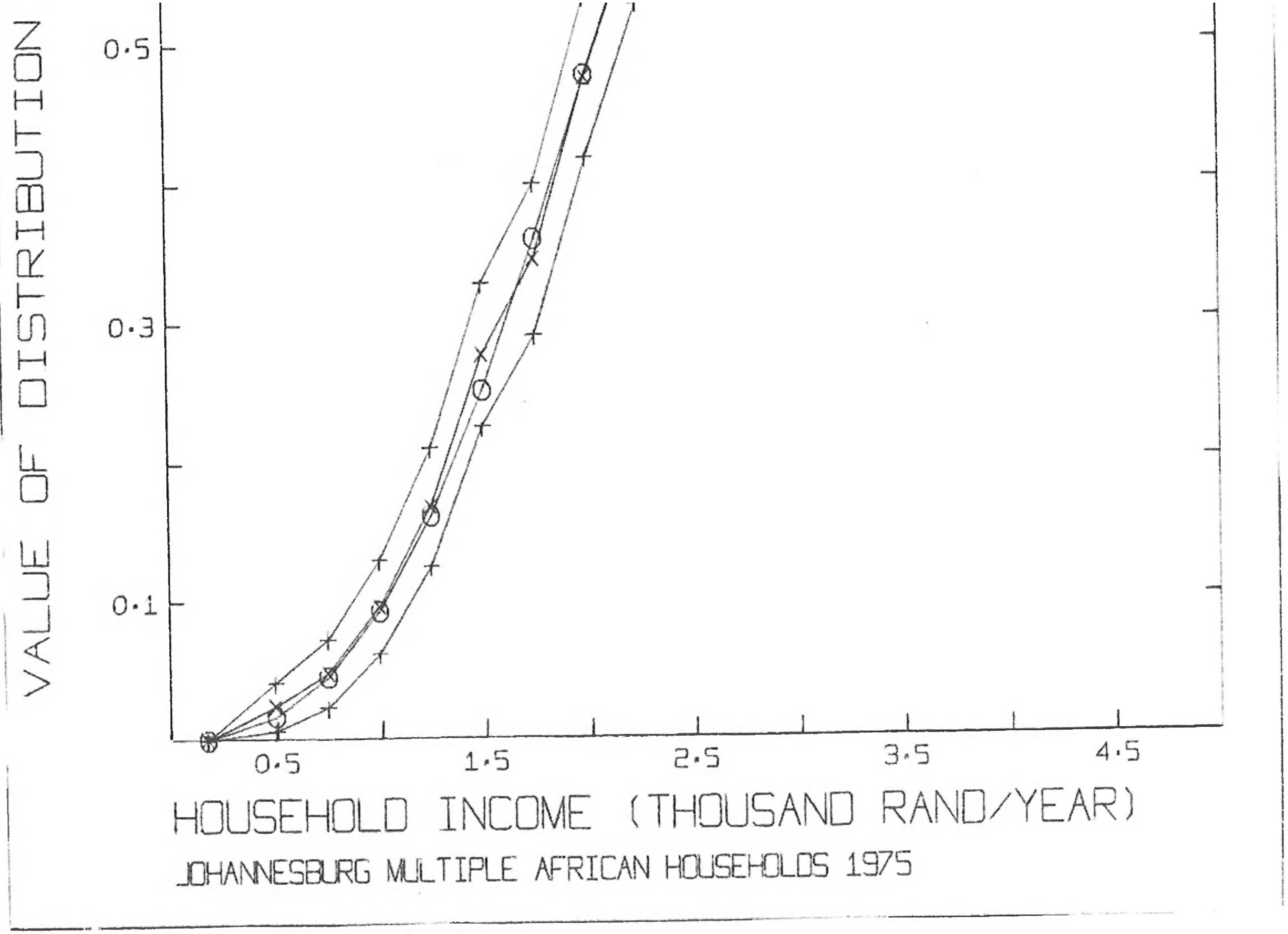

VALUE OF DISTRIBUTION
0·1
0·3
0·5
0·5
1·5
2·5
3·5
4·5
HOUSEHOLD INCOME (THOUSAND RAND/YEAR)
JOHANNESBURG MULTIPLE AFRICAN HOUSEHOLDS 1975

FIGURE 5.

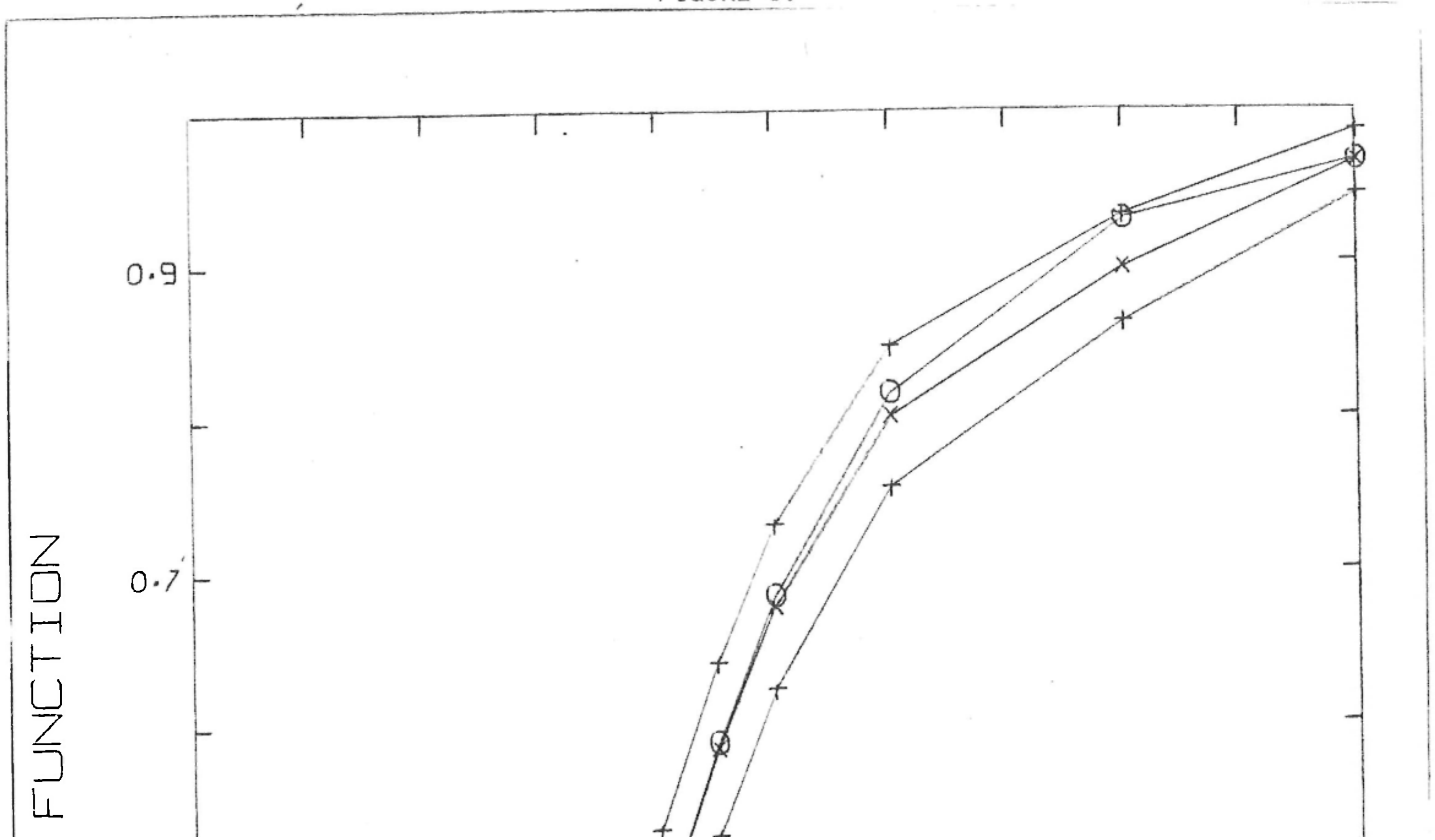
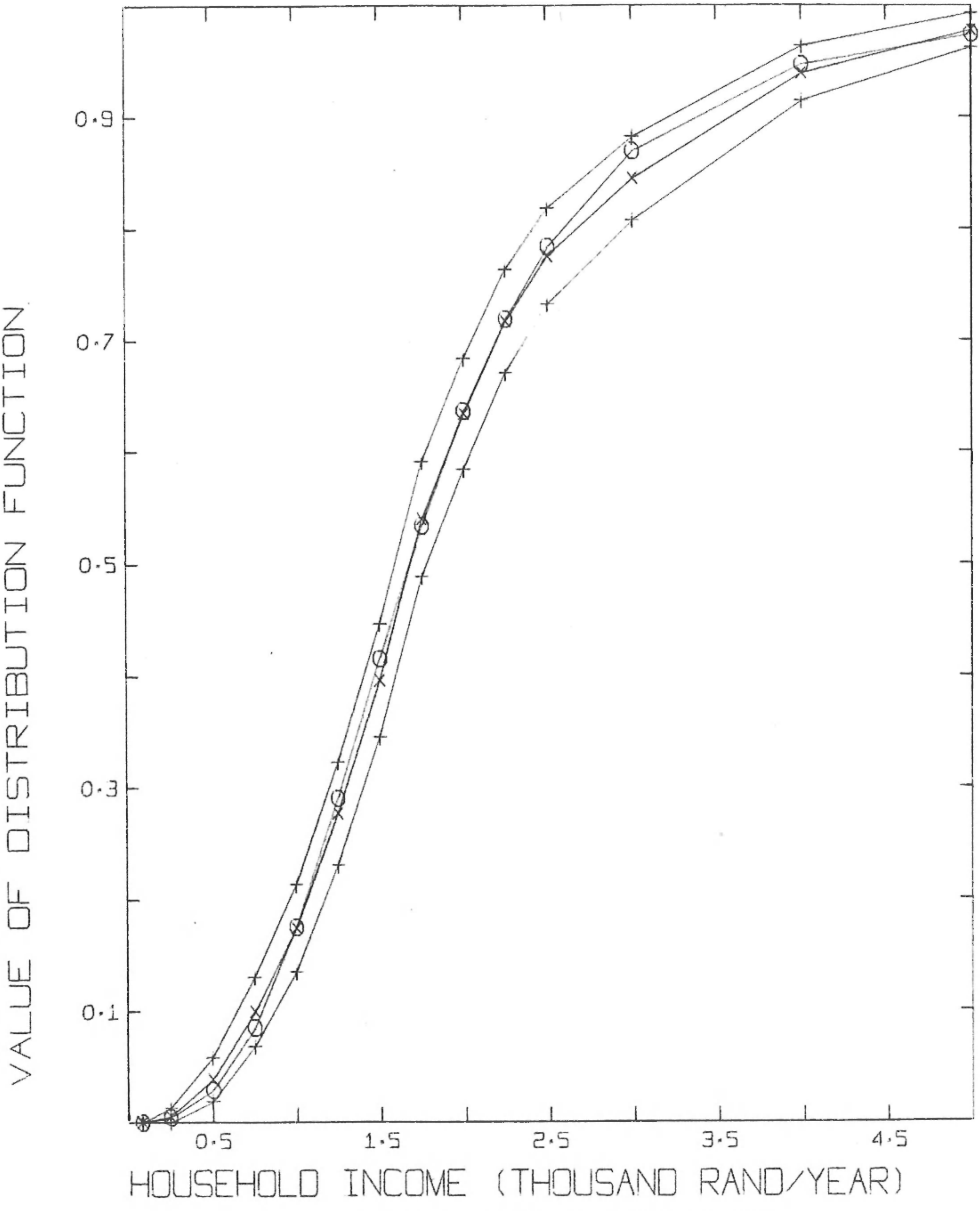
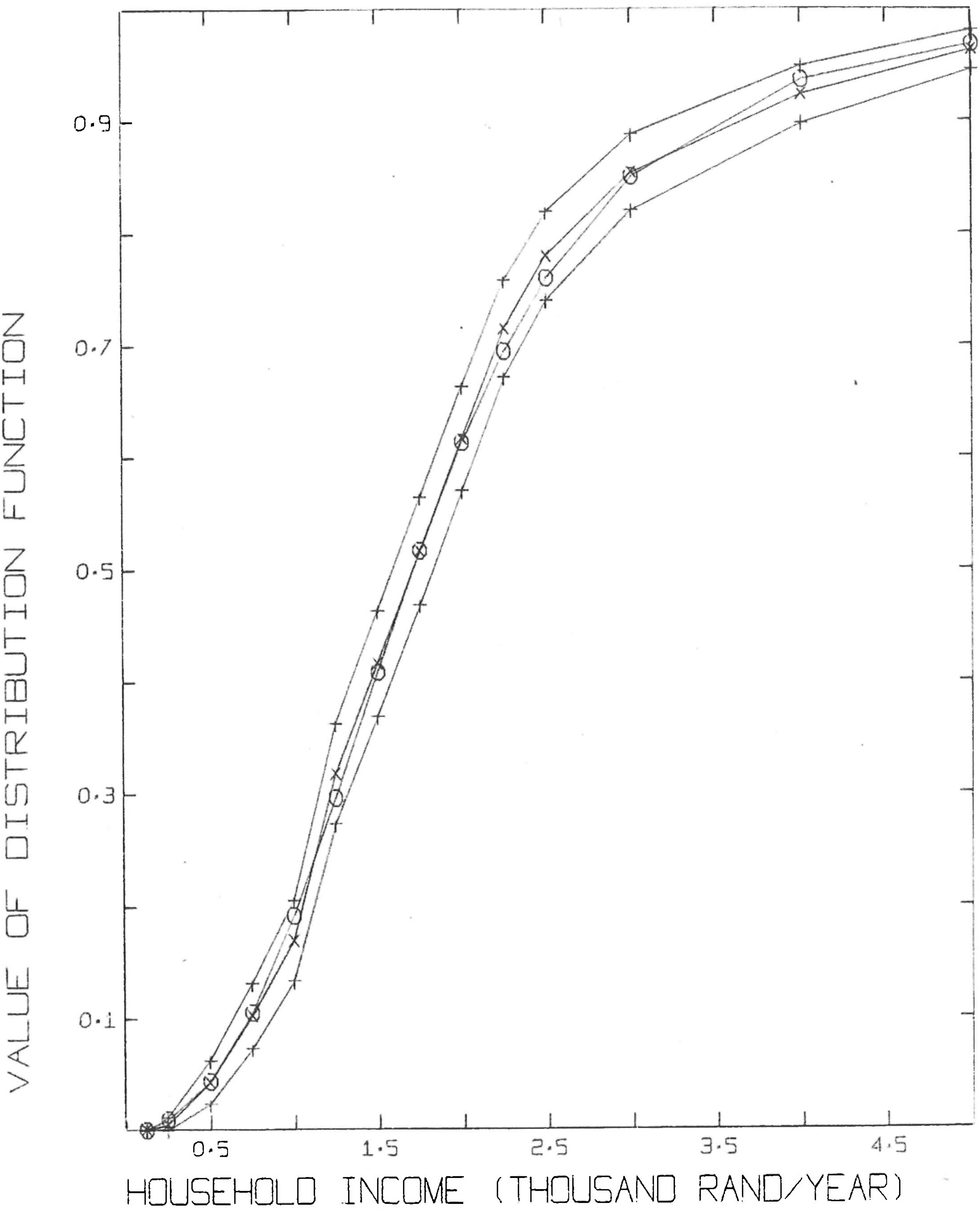


FIGURE 6.



PORT ELIZABETH MULTIPLE AFRICAN HOUSEHOLDS 1975

FIGURE 7.



PRETORIA MULTIPLE AFRICAN HOUSEHOLDS 1975

TABLE V.

GOODNESS OF FIT OF CHAMPERNOWNE DISTRIBUTION FUNCTION. (1975)

City	χ^2	degrees of freedom	χ^2_c	Sample size
Bloemfontein	9,07	9	16,92	293
Cape Town	4,66	11	19,68	196
Durban	12,08	12	21,03	391
East London	10,95	12	21,03	288
Johannesburg	14,62	11	19,68	292
Port Elizabeth	10,12	12	21,03	361
Pretoria	15,21	12	21,03	418

Thus, the hypothesis that the Champernowne curve fits the data is not rejected at the 5% level of significance for any of the cities. The power of the test is relatively low, i.e. the probability of accepting the curves as good fits when they are not is relatively high, because of small sample sizes; it is likely that the hypothesis would be rejected in some cases if the samples were, say, 10 or 100 times as large as they are.

Table VI sets out the various estimates of the general inequality measures i_1 to i_6.

<u>TABLE VI</u>

GENERAL INEQUALITY MEASURES 1975.

Measure	Cubic spline	Champernowne	Lower grouping limit	Upper grouping limit	Lower 95% conf. limit	Upper 95% conf. limit	Cubic spline	Champernowne	Lower grouping limit	Upper grouping limit	Lower 95% conf. limit	Upper 95% conf. limit
	BLOEMFONTEIN.						CAPE TOWN.					
i_1	0,268	0,322	0,255	0,263	0,207	0,328	0,234	0,281	0,240	0,247	0,170	0,299
i_2	0,257	0,237	0,255	*	0,224	0,288	0,296	0,290	0,282	*	0,251	0,335
i_3	0,146	0,139	0,143	0,186			0,159	0,165	0,157	0,177		
i_4	0,320	0,289	0,305	*			0,394	0,358	0,338	*		
i_5	0,291		0,286	0,293	0,259	0,323	0,306		0,303	0,307	0,266	0,346
i_6	0,141	0,143	0,136	0,143			0,138	0,153	0,140	0,144		
	DURBAN.						EAST LONDON.					
i_1	0,229	0,263	0,232	0,252	0,184	0,274	0,282	0,313	0,323	*	0,219	0,345
i_2	0,239	0,236	0,240	0,250	0,213	0,264	0,318	0,318	0,306	*	0,280	0,352
i_3	0,134	0,137	0,136	0,142			0,185	0,190	0,187	0,209		
i_4	0,276	0,276	0,275	0,293			0,393	0,393	0,359	*		
i_5	0,289		0,286	0,291	0,253	0,315	0,345		0,342	0,346	0,307	0,383
i_6	0,126	0,134	0,128	0,136			0,167	0,177	0,180	0,183		

* Upper limits not well calculated by slightly oversimplified programme.

TABLE VI - continued

GENERAL INEQUALITY MEASURES 1975.

Measure	Cubic spline	Champer-nowne	Lower grouping limit	Upper grouping limit	Lower 95% conf. limit	Upper 95% conf. limit	Cubic spline	Champer-nowne	Lower grouping limit	Upper grouping limit	Lower 95% conf. limit	Upper 95% conf. limit
	JOHANNESBURG.						PIETERMARITZBURG.					
i_1	0,202	0,267	0,224	0,228	0,157	0,248	0,264	-	0,297	0,311	0,204	0,323
i_2	0,217	0,214	0,231	0,241	0,188	0,244	0,246	-	0,259	0,266	0,214	0,275
i_3	0,119	0,124	0,130	0,135			0,150	-	0,165	0,170		
i_4	0,237	0,251	0,265	0,281			0,276	-	0,293	0,303		
i_5	0,281		0,279	0,283	0,253	0,309	0,335		0,332	0,337	0,299	0,371
i_6	0,118	0,127	0,123	0,126			0,148	-	0,166	0,173		
	PORT ELIZABETH.						PRETORIA.					
i_1	0,249	0,312	0,250	0,260	0,199	0,300	0,273	0,341	0,312	0,317	0,221	0,324
i_2	0,264	0,255	0,259	0,273	0,234	0,292	0,276	0,296	0,274	0,285	0,248	0,302
i_3	0,152	0,154	0,149	0,156			0,163	0,179	0,171	0,176		
i_4	0,307	0,300	0,300	0,326			0,320	0,361	0,317	0,336		
i_5	0,305		0,302	0,306	0,275	0,335	0,332		0,329	0,333	0,302	0,362
i_6	0,141	0,156	0,140	0,146			0,155	0,175	0,170	0,174		

* Upper limits not well calculated by slightly oversimplified programme.

The following should be noted from Table VI:

(a) in general, the range between the grouping limits is considerably smaller than the range between the 95% confidence limits i.e. the sampling error is greater than the grouping error. So great is the sampling error, in fact, that it would take an enormous shift in the estimates of the general inequality measures for us to be certain that a real change had taken place. This follows from the small sample size and is a major limitation imposed by the data source.

(b) in all cases, the Champernowne estimate of i_2 lies within the 95% confidence limits; in three cases, however, the Champernowne estimate of i_1 lies outside these limits. In all three cases, the Champernowne estimate lies above the upper limit. This is probably because of the slightly unsatisfactory treatment of the tails (particularly the upper tail) by the spline function.

Table VII contains the results of two tests for consistency of the estimates of i_1, i_2, i_3, i_4 and i_6 from the cubic spline and from the Champernowne function. Medians for seven cities are compared and Spearman rank correlation coefficients calculated.

TABLE VII.

CONSISTENCY TESTS FOR GENERAL INEQUALITY MEASURES, 1975.

Measure	Spearman rank	Median	
		cubic spline	Champernowne
i_1	0,857	0,249	0,312
i_2	0,964	0,264	0,255
i_3	1,000	0,152	0,154
i_4	0,857	0,320	0,300
i_6	0,964	0,141	0,153

All the Spearman rank correlation coefficients are significant at the 5% level and, with the exception of i_1 (for reasons already discussed), the medians are in close agreement. So the two methods of obtaining point estimates for the general inequality measures are consistent, at least when it comes to dealing with the data being studied here.

Patterns of inequality.

Despite the large sampling errors, one might be able to build up some sort of idea of what has been happening to relative inequality by comparing all the measures, obtained from the cubic spline and Champernowne functions, for each city where surveys were taken in both 1970 and 1975. Table VIII sets out the results.

TABLE VIII.

COMPARISONS OF GENERAL INEQUALITY MEASURES, 1970 AND 1975.

	Cape Town		Durban		East London	
	Cubic spline	Champer-nowne	Cubic spline	Champer-nowne	Cubic spline	Chamer-nowne
i_1	++	-	+	--	-	+
i_2	++	++	0	-	0	0
i_3	++	++	+	-	0	0
i_4	++	++	-	-	0	0
i_5	+		0		0	
i_6	++	+	+	-	0	+
Total	11 +	6 +	2 +	6 -	1 -	2 +

	Johannesburg		Port Elizabeth		Pretoria	
	Cubic spline	Champer-nowne	Cubic Spline	Champer-nowne	Cubic spline	Champer-nowne
i_1	-	+	-	--	+	0
i_2	-	0	-	-	++	++
i_3	-	0	-	-	++	++
i_4	--	0	--	-	++	++
i_5	0		0		++	
i_6	0	0	0	-	++	++
Total	5 -	1 +	5 -	6 -	11 +	8 +

Key: ++/-- an increase/decrease of at least 20% from 1970 to 1975.
+/- an increase/decrease of 5-20% from 1970 to 1975.
0 a change of less than 5% from 1970 to 1975.

To speak more loosely than is really desirable, Table VIII clearly suggests rising relative inequality in Cape Town and Pretoria and falling inequality in Port Elizabeth. In East London there has been no change in relative inequality to speak of, while in Durban and Johannesburg, the situation is unclear, one set of comparisons in each case indicating falling inequality and the other, if anything, slightly rising inequality.

Testing for systematic variation of Champernowne measures with each other and the mean over the 1975 cross-section is carried out by calculating Spearman rank correlation coefficients between each pair of the specialist inequality measures j_1, j_2, j_3, the scale parameter and mean. Results are set out in Table IX.

TABLE IX.

SPEARMAN RANK CORRELATION COEFFICIENTS BETWEEN EACH PAIR OF j_1, j_2, j_3, x_0 AND $\bar{x}$, 1975.

	j_1	j_2	j_3	x_0	$\bar{x}$
j_1	-	-0,821*	-0,321	-0,795*	-0,661
j_2		-	0,143	0,527	-0,214
j_3			-	0,491	-0,286
x_0				-	0,348
$\bar{x}$					-

Note: * denotes coefficients significant at the 5% level.

Only the correlations between j_1 and j_2 and j_1 and x_0 are significant; i.e. α- inequality decreases as β- inequality increases and α- inequality decreases as the scale parameter increases. γ- inequality is not related significantly to either α- or β- inequality on the one hand or the scale parameter or mean income on the other.

Comparisons between j_1, j_2 and j_3 in 1970 and 1975 can be carried out for six cities and the results are set out in Table X.

TABLE X.

COMPARISONS OF j_1, j_2, j_3 IN 1970 AND 1975.

	j_1			j_2			j_3		
	1970	1975	Inc(+)/ Dec(-)	1970	1975	Inc/ Dec	1970	1975	Inc/ Dec
Cape Town	0,351	0,239	-	0,123	0,429	+	0,262	0,325	+
Durban	0,334	0,256	-	0,280	0,419	+	0,275	0,273	-
East London	0,190	0,183	-	0,928	1,512	+	0,337	0,325	-
Johannesburg	0,235	0,295	+	0,600	0,222	-	0,243	0,277	+
Port Elizabeth	0,369	0,282	-	0,187	0,488	+	0,326	0,267	-
Pretoria	0,340	0,289	-	0,224	0,421	+	0,230	0,309	+

α- inequality appears to have decreased while β- inequality has increased, except in the case of Johannesburg. The time-series results therefore show the same inverse relationship between α- and β- inequality as the cross-section study. There is no clear trend in γ- inequality.

The poverty analysis.

In order to assess how the situation as regards households below the MLL has evolved between 1970 and 1975, we tabulate the parameters H, I, G and P as well as median incomes (in 1975 prices) for each of the six cities for which survey results in both years are available.[30]

TABLE XI

H, I, G, P AND $\bar{x}$ IN 1970 AND 1975.

	H			I			G			P			$\bar{x}$		
	70	75	Inc/ Dec	70	75	Inc/ Dec	70	75	Inc/ Dec	70	75	Inc/ Dec	70	75	Growt (% p.a
Cape Town	0,295	0,273	-	0,263	0,350	+	0,151	0,212	+	0,111	0,133	+	1,581	1,856	3,3
Durban	0,501	0,196	-	0,308	0,277	-	0,182	0,161	-	0,218	0,077	-	1,252	1,902	8,7
East London	0,580	0,352	-	0,373	0,331	-	0,229	0,198	-	0,300	0,163	-	0,992	1,378	6,8
Johannesburg	0,340	0,151	-	0,309	0,335	+	0,183	0,201	+	0,148	0,071	-	1,459	2.047	7,0
Port Elizabeth	0,450	0,257	-	0,417	0,284	-	0,264	0,166	-	0,257	0,103	-	1,230	1,675	6,4
Pretoria	0,443	0,280	-	0,274	0,277	+	0,159	0,160	+	0,173	0,110	-	1,266	1,709	6,2

In all cases, then, the head-count measure has declined; in three out of six cases I and G have declined and in the other three, they have increased. The decrease in the proportion of households below the MLL serves to outweigh the increase in I and G in two of three cases, when it comes to the composite measure P; In Cape Town where the median income has grown at a markedly slower rate that elsewhere, the increase in I and G outweigh the decrease in H and P has risen.

γ' is, of course, related to γ; since γ has not varied much over time or over the 1975 cross-section (and from, Table X, it appears that the same is true for γ'), the following analysis is introduced to relate H and P to $\tilde{x}$ and p (the MLL). Suppose that for each city in 1970 and 1975, the distribution function F(x) can be represented by

$$F(x) \quad = \quad Ax^{\gamma''} \qquad 0 < x < \tilde{x} \qquad \text{where } \gamma'' \text{ is a constant}$$

This amounts to introducing the hypothesis that the pseudo-Pareto law holds up to median incomes and that the pseudo-Pareto constant is the same for all cities in both years.

Now $F(\tilde{x}) = \frac{1}{2}$ since $\tilde{x}$ is the median

$$A\tilde{x}^{\gamma''} = \tfrac{1}{2} \text{ or } A = \frac{1}{2\tilde{x}^{\gamma''}}$$

$$H = F(p) = \frac{p^{\gamma''}}{2\tilde{x}^{\gamma''}} \qquad \text{--- (1)}$$

$$2H = \left(\frac{\tilde{x}}{p}\right)^{-\gamma''}$$

$$\underline{\text{or } \ln 2H = -\gamma'' \ln \left(\frac{\tilde{x}}{p}\right)}$$

i.e. if the hypothesis is true and we regress ln 2H against $\ln(\frac{\tilde{x}}{p})$, we should find a straight line through the origin with slope $-\gamma''$. Table XII contains the values of H and $\frac{\tilde{x}}{p}$ used in the regression.

TABLE XII

VALUES OF H AND $\frac{\tilde{x}}{p}$ IN 1970 AND 1975.

	1970		1975	
	H	$\frac{\tilde{x}}{p}$	H	$\frac{\tilde{x}}{p}$
Bloemfontein			0,547	0,942
Cape Town	0,295	1,228	0,273	1,442
Durban	0,501	0,998	0,196	1,517
East London	0,580	0,895	0,352	1,244
Johannesburg	0,340	1,213	0,151	1,702
Pietermaritzburg			0,175	1,583
Port Elizabeth	0,450	1,060	0,257	1,391
East London	0,443	1,066	0,281	1,439

from which we find:

$$\ln 2H = -\,2{,}058 \ln \left(\frac{\tilde{x}}{p}\right) \qquad R^2 = 0{,}954$$

(24,6) (t- value in brackets)

i.e. $\underline{\gamma'' = 2{,}058}$

Values of $\frac{\tilde{x}}{p}$ (the ratio of the median income to the MLL) required to achieve various values of H are tabulated in Table XIII. The percentage increases in $\tilde{x}$ between successive targets are also set out, showing how it gets progressively harder to reduce H.

TABLE XIII

REQUIRED VALUES OF $\frac{\tilde{x}}{p}$ FOR TARGET VALUES OF H.

H	$\frac{\tilde{x}}{p}$	$\Delta\tilde{x}$ (%)
0,30	1,282	-
0,25	1,400	9,2
0,20	1,561	11,5
0,15	1,795	15,0
0,10	2,186	21,8
0,05	3,061	40,0

It follows from (1), however, that $\frac{\Delta H}{H} = -\gamma'' \frac{\Delta \tilde{x}}{\tilde{x}}$, so that proportional increases in the median income will cause a constant proportional reduction in H, no matter what the initial levels of $\tilde{x}$ and H are.

Since $P = H \frac{3\gamma''+1}{(\gamma''+1)(2\gamma''+1)}$,P is proportional to H if γ'' is constant, the constant of proportionality being 0,459.

It remains to report the results of the simulation study. The distributions of households by number of earners and by household income generated by the programme were compared with the sample tabulations and χ^2 goodness-of-fit tests were carried out. The results are set out in Table XIV.

TABLE XIV.

GOODNESS-OF-FIT OF SIMULATED DISTRIBUTIONS OF HOUSEHOLDS BY EARNERS AND BY HOUSEHOLD INCOME

City	Distribution by earners				Distribution by household income.			
	χ^2	degrees of freedom	χ^2_c (5%)	χ^2_c (1%)	χ^2	degrees of freedom	χ^2_c (5%)	χ^2_c (1%)
Cape Town	1,16	3	7,82	11,34	12,44	11	19,68	24,72
Durban	7,51*	2	5,99	9,21	20,11	12	21,03	26,22
East London	1,50	2	5,99	9,21	20,00*	11	19,68	24,72
Johannesburg	6,86*	2	5,99	9,21	35,45**	12	21,03	26,22
Port Elizabeth	6,90	3	7,82	11,34	23,98*	12	21,03	26,22
Pretoria	13,72**	3	7,82	11,34	15,42	12	21,03	26,22

* significant at 5% level
** significant at 1% level

Goodness-of-fit is generally not too bad, four out of six cities investigated having at most one χ^2- test significant at the 5% but not at the 1%, level. The values of H and P obtained for these cities and the percentage differences between these values and those reported in Table XI are displayed in Table XV.

TABLE XV.

VALUES OF H AND P OBTAINED FOR FOUR CITIES IN 1975 BY SIMULATION.

City	H		P	
	Simulation value	Percentage difference from Table XI	Simulation value	Percentage difference from Table XI
Cape Town	0,231	- 15%	0,113	- 15%
Durban	0,251	+ 28%	0,131	+ 70%
East London	0,314	- 11%	0,195	+ 20%
Port Elizabeth	0,252	- 2%	0,130	+ 26%
Median - Table XI values	0,265		0,118	
Median - Simulation	0,251	- 5%	0,131	+ 11%

The variations are fairly large, especially in the case of the Sen poverty measure. Unfortunately, no 'true' values are available to enable one to assess which procedure is the more accurate.

The wage increase simulations allow us to estimate the size of the increase necessary to effect a (say) 10% drop in H and P respectively. This can be compared with the estimate produced by the regression equation (the same for H and P, since in that case, P and H are proportional). The results appear in Table XVI.

TABLE XVI.

ACROSS-THE-BOARD WAGE INCREASES REQUIRED TO ACHIEVE A 10% REDUCTION IN H AND P IN FOUR CITIES IN 1975 (PERCENT)

City	Simulation		Regression
	10% reduction in H	10% reduction in P	10% reduction in H and P
Cape Town	5,4	6,8	5,2
Durban	6,2	5,2	5,3
East London	8,8	9,4	5,2
Port Elizabeth	8,0	7,7	5,2

The estimates are reasonably close in the cases of Durban and Cape Town, but further apart in East London and Port Elizabeth.

Perhaps the most interesting results are the estimates of the critical wage elasticities of demand for labour. They are reported in Table XVII.

TABLE XVII.

CRITICAL WAGE ELASTICITIES OF DEMAND FOR LABOUR IN FOUR CITIES IN 1975.

City	H		P	
	Unemployment among 2nd and subsequent earners	Unemployment among all earners	Unemployment among 2nd and subsequent earners	Unemployment among all earners
Cape Town	-	0,54	-	0,18
Durban	1,72	0,69	2,33	0,27
East London	1,25	0,99	1,12	0,27
Port Elizabeth	0,79	0,34	0,86	0,17

It turns out that estimates of critical elasticities are very sensitive to the assumptions made about the incidence of unemployment. Much lower critical elasticities are found (both in the case of H and P) if first earners become unemployed. On the assumptions made, if first earners become unemployed, the elasticity of demand for labour is likely to be high enough to make self-defeating attempts to reduce P by across-the-board proportional wage increases; and decreases in H arising from such wage increases will be substantially offset by the corresponding drop in employment.

Three features of what may actually happen would serve to raise the critical elasticities somewhat in this case:

- the programme sets the income of an unemployed person to zero; this is probably too dire, as some such will receive unemployment insurance or manage to obtain a little income from casual work or informal sector activities
- a household with no earners may not be able to pay rent, once savings are run down. Eviction will follow and the household will disappear from the universe being considered (settled African households) and H and P among those that remain will rise
- a household with no earners may merge with another household (containing relatives) with an uncertain effect on H but probably reducing P.

If first earners do not become unemployed, then it is likely that the unemployment effect will not reduce very much the drop in H and P resulting from wage increases.

It should be stressed that this analysis takes place within a static framework. In a dynamic context, it appears that it is quite possible to reduce H and P substantially as real wages rise - comparisons between 1970 and 1975 demonstrate this.

Scope of this study

A few observations are required to indicate the limitations of the scope of this study. The universe from which the samples were drawn is tabulated in Table XVII.

TABLE XVII.

ESTIMATED SETTLED AFRICAN POPULATIONS IN CITIES STUDIED, 1970 AND 1975.

	1970	1975	Growth rate (% p.a.)
Bloemfontein		64 894	
Cape Town	89 880	65 194	-6,2
Durban	238 832	445 900	13,3
- Duncan Village	37 517	39 900	1,2
East London - Mdantsane		100 100	
Johannesburg	578 454	761 482	5,7
Pietermaritzburg		98 000	
Port Elizabeth	172 763	233 471	6,2
Pretoria	237 191	226 921	-1,1
Total - areas incl. in 70.	1 354 637	1 772 868	5,5
Total - areas incl. in 75.		2 035 952	
TOTAL AFRICAN POPULATION.	15 918 000	18 173 000	2,7
% - areas incl. in 70.	8,5	9,8	
% - areas incl. in 75.		11,2	

Sources: BMR reports
Department of Statistics, Statistical News Release P 11 (25.10.76).

In other words, only about one-twelfth of the African population was included in the 1970 surveys discussed here and one-ninth in the 1975 surveys. Excluded from consideration here are:

- settled families in other metropolitan areas (notably the East and West Rand and Vereeniging) where conditions may be comparable to those found in this study
- families living as squatters rather than in settled housing in the metropolitan areas.

- settled families in smaller towns and border areas
- settled families on 'white farms' and in the 'homelands'
- migrants and their dependents.

We cannot apply the conclusions of this study to the last four categories; in each conditions are likely to be worse than those found here.

SUMMARY OF CONCLUSIONS:

The following are the principal conclusions of this study:

(a) A cubic spline function, designed to reproduce the values of the observed distribution function at base points, and a Champernowne density function fitted to the tails, the mean, the median and the geometric mean of the observed distribution, yield broadly consistent estimates for five general inequality measures (the normalised coefficient of variation, the normalised coefficient of variance of logarithms, the Atkinson index with inequality aversion parameters of one and two, and the normalised Theil measure). The Champernowne distribution function is accepted as fitting the data at the 5% level of significance on a χ^2 goodness-of-fit test.

(b) Small sample sizes set the limits to the accuracy of the results obtained. The sampling error in the normalised coefficient of variation and the normalised coefficient of variance of logarithms is large and that in the Gini coefficient is moderate; the sampling error is generally considerably greater than the error arising from the grouping of data into income classes.

The estimates of α are based on the incomes of very few households and are thus rather unreliable. It is also possible that the Champernowne distribution function would fail the goodness-of-fit test were samples bigger.

(c) Quite low values are found for the general measures of inequality among settled African households. Values of α are probably high compared with those found in many countries, indicating little inequality among the relatively rich.

(d) Comparisons of general inequality measures in 1970 and 1975 yield the conclusions that relative inequality has increased in two cities (Cape Town and Pretoria), decreased in one (Port Elizabeth) and remained virtually constant in one (East London). Results for Durban and Johannesburg are less clear - if anything, relative inequality has decreased in both places.

(e) Cross-section and time series studies both yield the conclusion that α- inequality is inversely related to β- inequality (inequality among the middle incomes) in urban multiple African households; the indications are that β- inequality has increased and α- inequality decreased (except in Johannesburg) between 1970 and 1975. The explanation for this possibly lies in the rapid raising of wages among unskilled and semi-skilled workers in industrial and commercial employment and in the opening up of more levels of employment to Africans over the period.

(f) No trends can be found in γ- inequality (inequality among the poor). γ is generally just above two and there is little dispersion in the values found. γ- inequality, in addition to being related to conditions in the labour market and demographic circumstances, must also be related to the cost of and availability of housing in the areas studied. Poor people are much more heavily represented in squatter areas than in townships;[31] the degree to which this is so depends on the supply of low-cost housing in relation to demand.

(g) Progress (rapid in some cities, less so in others) has been made reducing the proportion of households below the Minimum Living Level and the Sen poverty measure between 1970 and 1975. It will become progressively more difficult to reduce the proportion further. By contrast, there has been no trend in the levels of the average income-gap (the proportion of the MLL by which these households fall short of it) or in the Gini coefficient associated with them.

(h) If, within a static framework, we consider the raising of wages proportionately across the board as a way of reducing poverty, the effectiveness of such a move depends on the incidence and extent of the resulting drop in employment. If first earners do not become unemployed, the measure is likely to be effective; if they do, it will be much less so, and possibly self-defeating. Other factors, principally the relative rates of growth of urban population and employment, operate in a dynamic framework to complicate the picture.

(i) The conclusions of this study apply to about one-ninth of the total African population in South Africa. Conditions may be broadly the same amongst a further small proportion located in housing estates in major urban areas not covered by the surveys. It is certain that the proportion of poor would be increased if squatter households in the metropolitan areas, households in smaller towns, rural households and migrants and their dependents were to be considered. It would be a mistake, therefore, to generalise the results of this study to the African population as a whole.

NOTES.

1 Acknowledgements: I am grateful to Michael McGrath, senior lecturer in the Department of Economics, University of Natal, Durban, for introducing me to the contemporary literature on the size distribution of income and for the many discussions we have had on its application to South African problems. Valuable assistance with computing was received from David Wallis and Clive Reid of the Computer Centre, University of Natal, Pietermaritzburg. Norman Bromberger of the DSRG provided useful comments on an earlier draft. Naturally, any mistakes which remain are my own responsibility.

2 On this, see (i) A. Spandau, Income distribution and economic growth in South Africa, unpublished D. Com. thesis, University of South Africa, 1971.
(ii) M.D. McGrath, Racial income distribution in South Africa, University of Natal, Black-White Income Gap Research Report no.2, University of Natal, Durban, 1977.
(iii) J. Nattrass, The narrowing of wage differentials in South Africa, South African Journal of Economics, vol. 45 no.4, Dec.1977, pp. 408-432.

3 Bureau of Market Research: Reports nos. 27.1, 27.2, 27.4, 27.7, 27.10 and 27.11 (1970 surveys) and nos. 50.1, 50.2, 50.3, 50.4, 50.10, 50.11, 50.12 and 50.13 (1975 surveys). In 1970, surveys were also carried out in Krugersdorp and Tembisa; they are not considered here as these areas were not studied again in 1975.

4 I am equating 'settled' with 'multiple' African households; 'single' African households consist overwhelmingly of migrant workers and domestic servants living on premises owned by members of other races. The families of such persons live in rural areas; in the case of 'multiple' households, at least the nucleus of a family lives in town. It seems sensible, therefore, to reserve the surveys of 'single' households for a somewhat different study.

5 See especially D.G. Champernowne : A comparison of measures of inequality of income distribution; Economic Journal, vol.84.(1974) pp. 787-816.

6 The Minimum Living Level in a number of centres for February and August each year is calculated by the Bureau of Market Research. See, for example, report No.47. The minimum and supplemented living levels of non-whites residing in the main and other selected urban areas of the Republic of South Africa, August 1975 by M. Loubser. Opinions on the minimum income required for subsistence in the short and long runs differ; The Institute for Planning Research at the University of Port Elizabeth publishes more generous estimates, entitled the 'household subsistence level' and 'household effective level' respectively. All these measures (and others) are comprehensively discussed in P.A. Ellison, P.N. Pillay and G.G. Maasdorp: The 'poverty datum line' debate in South Africa : an appraisal, Department of Economics, University of Natal, Durban, 1975.

7 On this, see S. Kuznets, Demographic aspects of the size distribution of income: an exploratory essay, Economic Development and Cultural Change, vol.25. no.1, Oct. 1976 pp. 1-94.

8 A general definition of personal (and, by extension, household) income is provided by H.C. Simons in ed. Baker and Harcourt: Readings in the concept and measurement of income.

"Its calculation implies estimate of the amount by which the value of a person's store of property rights would have increased, as between the beginning and end of the period, if he had consumed (destroyed) nothing".

(p. 68). This definition permits negative incomes; the Bureau of Market Research technique of measurement , however, excludes consideration of capital losses (although allowing for 'income from property'), so that all its observations of household income are non-negative.

9 see H.J. Larsen: Introduction to probability theory and statistical inference, Second Edition, Wiley, 1974.

10 see B. Carnahan and J.O. Wilkes: Digital computing and numerical methods, Wiley, 1973 pp. 307-310.

11 These measures are selected as they are all discussed in Champernowne op. cit.

12 See A.B. Atkinson, On the Measurement of Inequality, Journal of Economic Theory, vol.2, pp. 244-263, 1970.

13 On this see H. Theil, Economics and Information Theory, North Holland Publishing Company, 1967, pp. 128-134.

14 For the lognormal distribution see Aitcheson and Brown: The lognormal distribution, Cambridge University Press, 1959, p.39; and for the Gini coefficient, see Kendall and Stuart: The advanced theory of statistics, vol.1. (second edition), Charles Griffin, 1963, pp. 241-242.

15 Champernowne op. cit. p. 801. A companion function exists:

$$f(x) = \frac{C'}{x\left[\left(\frac{x}{x_0}\right)^{-\gamma} + 2\cosh\beta\left(\frac{x}{x_0}\right)^{-\gamma^*} + \left(\frac{x}{x_0}\right)^{\alpha}\right]} \qquad x > 0$$

where
$$C' = \frac{\alpha^*\sinh\beta\,\sin\left(\frac{\gamma^*\pi}{\alpha^*}\right)}{\pi\sinh\left(\frac{\gamma^*\beta}{\alpha^*}\right)}$$

where the inequality among middle incomes is high.

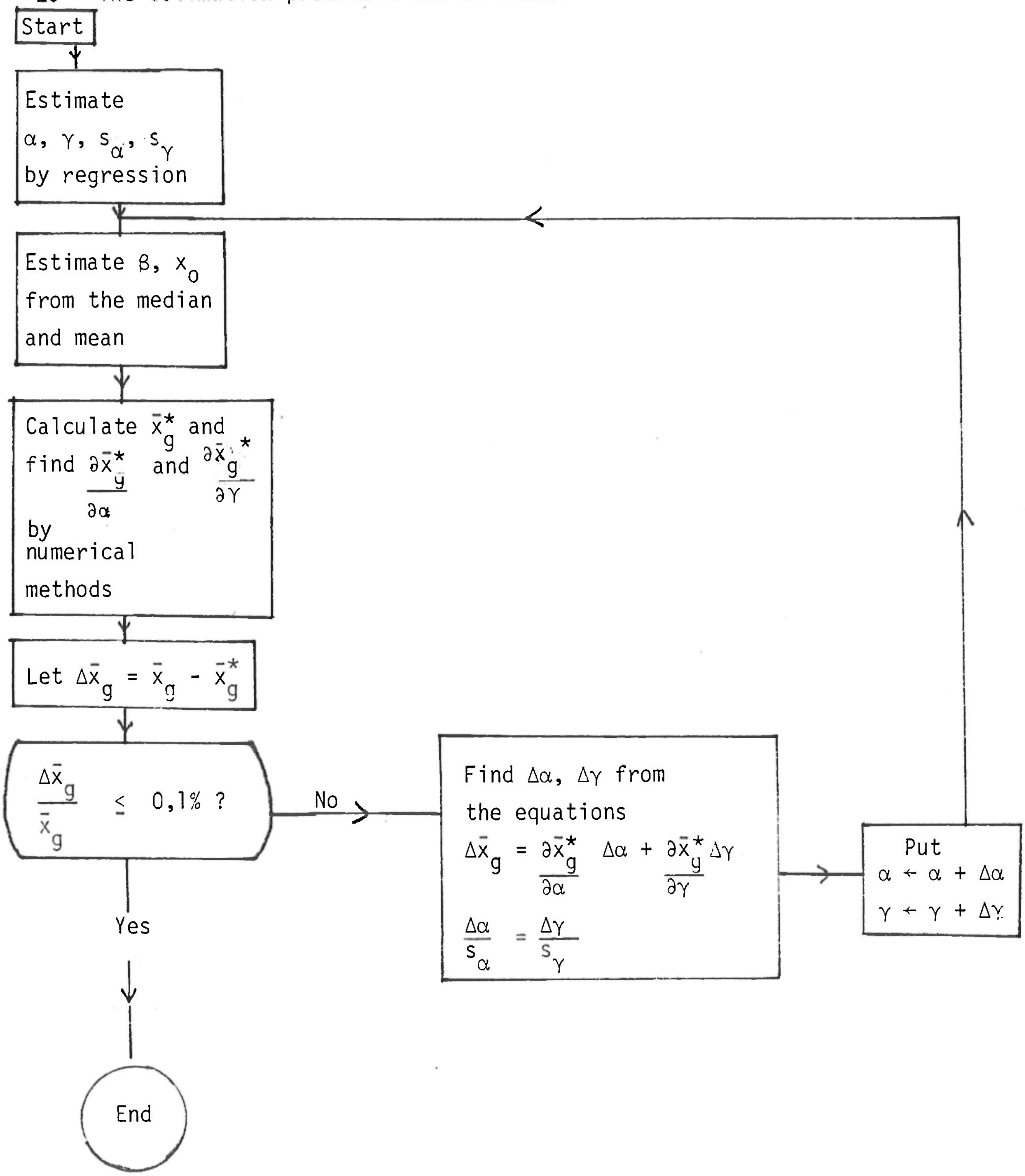
16 Champernowne mimeo notes equations 2.2.2 - 2.2.4. If cos β is replaced by cosh β, the trigonometric functions are replaced by the corresponding hyperbolic functions in the definition of β_t and b_t.

17 Ibid. equation 3.2.9. Champernowne measures x in terms of x_0 throughout. If x_0 is to be estimated, x must be measured in terms of other units in the first instance and then reduced to a multiple of x_0 by division by that quantity.

18 In fact two intervals must be considered for $\beta \sim [0, \pi]$ first; if a solution exists in that interval then we use the cos β form of the density function. If not, then we consider the interval $[0, 4{,}35]$ over which cosh β ranges from 1 to 39 i.e. j_2 from 0,5 to 10; if a solution exists in that interval, we use the cosh β form of the density function. If a solution cannot be found in either interval, we abandon the attempt to fit the Champernowne function.

19 Champernowne mimeo notes. equation 3.3.5.

20 The estimation procedure can be summarised in the form of a flowchart:



21 Champernowne op. cit. p. 793.

22 Champernowne mimeo notes equations 2.3.1 - 2.3.5 express these in terms of the auxiliary parameters:

$$i_1 = 1 - \frac{b_o\, p_1^2\, p_2}{p_o\, b_1^2\, p_2} \qquad i_2 = 1 - \frac{\gamma^{*2}}{\gamma^{*2} + p_o^2 - b_o^2}$$

$$i_3 = 1 - \frac{p_o\, b_1}{p_1\, b_o} \exp\left\{\frac{\beta_o - \pi_o}{\gamma^*}\right\} \qquad i_4 = 1 - \frac{b_{-1}\, p_o^2\, b_1}{p_{-1}\, b_o^2\, p_1}$$

$$i_6 = 1 - \frac{b_o\, p_1}{b_1\, p_o} \exp\left\{\frac{\pi_1 - \beta_1}{1 + \gamma^*}\right\}$$

23. A.K. Sen : Poverty: an ordinal approach to measurement, Econometrica vol.44. (Mar.1976) pp. 219 - 231.

24 Sen. op. cit. p. 223.

25 Ibid. p.227.

26 The Bureau of Market Research uses a rather extended definition of 'earner', which includes recipients of transfer incomes and which is such that virtually all household income is imputed to earners.

27 The binomial probability mass function P(x) is given by

$P(x) = \binom{n}{x} p^x (1-p)^{n-x}$ $x=0,1,2,\ldots n$. This had to be modified because the Bureau of Market Research never reports a household as having zero earners. The modification proceeds as follows:

Let $r = x+1$ and $N = n+1$

Then $P(r) = \binom{N-1}{r-1} p^{r-1}(1-p)^{N-r}$ $r=1,2,\ldots N$.

N, of course, is known (the household size) and p can be estimated as follows:

$\bar{r} = E(r)$, the expected value of r is the average number of earners per household. Now

$$E(r) = \sum_{r=1}^{N} rP(r) = \sum_{r=1}^{N} r\binom{N-1}{r-1} p^{r-1}(1-p)^{N-r}$$

$$= \sum_{x=o}^{n} (x+1)\binom{n}{x} p^x(1-p)^{n-x}$$

$$= \sum_{x=o}^{n} x\binom{n}{x} p^x(1-p)^{n-x} + \sum_{x=o}^{n} \binom{n}{x} p^x(1-p)^{n-x}$$

$$= np+1$$

$$p = \frac{\bar{r}-1}{n} = \frac{\bar{r}-1}{N-1}$$

28 Steps of this size are necessary to drown out simulation 'noise' i.e. to get a coefficient of $\frac{\Delta u}{u}$ significantly different from zero, at the 5% level.

29 The median standard error for α estimated by regression is 0,514 and for γ is 0,380; these are both quite large in comparison with median α and γ (14% and 17%, respectively) and, in the case of γ with the spread of observations, further indicating a lack in significant difference between estimates of γ for the different cities.

30 The MLL's used are those for the average household in each city; they are taken from BMR report no. 47 (see footnote 5). The mean MLL for all cities (R100,36 per month) is used as the MLL for Bloemfontein and Pietermaritzburg.

31 See my Socioeconomic characteristics of sixteen squatter communities on the Cape Peninsula in 1975, Southern Africa Labour and Development Research Unit (forthcoming).

Appendix 1

THE CUBIC SPLINE FUNCTION AND ITS MODIFICATIONS.

To simplify our task, we begin by transforming the extrapolation problem into an interpolation one by arbitrarily defining $x_{k+1} = 2x_k$ and assuming $f(x) = 0$ for $x > x_{k+1}$ i.e. for practical purposes, we assume no incomes fall above an upper limit of twice the lower bound of the open class. Then $F(x_{k+1}) = 1$; we also have $F(x_1) = 0$. Let $n = k+1$, then over successive intervals $[x_i, x_{i+1}]$ $i = 1,2,\ldots n-1$, we approximate $F(x)$ by a series of cubic polynomials $p_i(x)$. The coefficients of the polynomials are selected in such a way that

(i) $p_i(x)$ matches the known functional values at x_i and x_{i+1}

(ii) the first and second derivatives match each other at the intermediate bas points $x_2, \ldots, x_{n-1}$

(iii) the second derivatives of $p_1(x)$ and $p_{n-1}(x)$ at x_2 and x_n respectively are zero

This is achieved as follows:

Given n points $x_1 = 0, x_2, \ldots, x_n$ and values of $F(x_1), \ldots, F(x_n)$, we obtain $p_i(x)$ as follows:

Let $F_i = F(x_i)$ $i = 1, \ldots n$

$h_i = x_{i+1} - x_i$ $i=1,\ldots, n-1$

$\phi_i = p_i''(x_i)$ $i=1,\ldots, n$

We have $\phi_1 = \phi_n = 0$ and n-2 equations in n-2 unknowns of the form:

$$\frac{h_{i-1}}{h_i}\phi_{i-1} + 2\left(1+\frac{h_{i-1}}{h_i}\right)\phi_i + \phi_{i+1} = \frac{6}{h_i}\left(\frac{F_{i+1} - F_i}{h_i} - \frac{F_i - F_{i-1}}{h_{i-1}}\right)$$

$$i = 2,3,\ldots, n-1$$

Then $$p_i(x) = \frac{\phi_i}{6h_i}(x_{i+1} - x)^3 + \frac{\phi_{i+1}}{6h_i}(x-x_i)^3$$

$$+ \left(\frac{F_{i+1}}{h_i} - \frac{h_i\phi_{i+1}}{6}\right)(x-x_i) + \left(\frac{F_i}{h_i} - \frac{h_i\phi_i}{6}\right)(x_{i+1}-x) \quad i-1,2,\ldots n-1 \quad 1$$

The coefficients $p_i(x) = ax^3 + bx^2 + cx + d$ become:

a : $\dfrac{\phi_{i+1} - \phi_i}{6h_i}$

b : $\dfrac{\phi_i \, x_{i+1}}{2h_i} - \dfrac{\phi_{i+1} \, x_i}{2h_i}$

c : $\dfrac{\phi_{i+1} \, x_i^{\,2}}{2h_i} - \dfrac{\phi_i \, x_{i+1}^{\,2}}{2h_i} + \dfrac{F_{i+1}}{h_i} - \dfrac{h_i \phi_{i+1}}{6} - \dfrac{F_i}{h_i} + \dfrac{h_i \phi_i}{6}$

d : $\dfrac{\phi_i \, x_{i+1}^{\,3}}{6h_i} - \dfrac{\phi_{i+1} \, x_i^{\,3}}{6h_i} - \dfrac{F_{i+1} \, x_i}{h_i} + \dfrac{h_i \phi_{i+1} \, x_i}{6} + \dfrac{F_i \, x_{i+1}}{h_i} - \dfrac{h_i \phi_i \, x_{i+1}}{6}$

In other words, the spline function made up of the $p_i(x)$ is continuous and twice differentiable over the interval (x_1, x_n); it follows that an approximatation to the density function $f(x)$ is defined and is continuous over this interval. What does <u>not</u> follow is that $f(x) > 0$ over the entire interval as we require or that $f(x) = 0$ at the endpoints x_1 and x_n as as would be desirable. Since the spline function approximation to $f(x)$ (a quadratic) is most likely to become negative in the first and last classes (where we would expect the actual value to be small), modifications are introduced to $p_1(x)$ and $p_{n-1}(x)$ to prevent the approximation from becoming negative <u>and</u> to make $p_1'(x_1) = 0$ and $p'_{n-1}(x_n) = 0$; these modifications involve a re-estimation of x_n and a possible revision of x_1 to a value in the range $(0, x_2)$. Details of the modifications are discussed below:

Modifications

(a) First class : The condition $\phi_1 = 0$ is sacrificed but $F_1(0) = 0$ and $f_1(0) = 0$.

$$F_1(x) = ax^3 + bx^2 + cx + d$$
$$F_1(0) = d = 0$$
$$f_1(x) = 3ax^2 + 2bx + c$$
$$f_1(0) = c = 0$$
$$F_1(x) = ax^3 + bx^2$$
$$f_1(x) = 3ax^2 + 2bx$$

For continuity of the distribution and probability density functions at x_2 we require

$$ax_2^{\ 3} + bx_2^{\ 2} = F(x_2) = p_2(x_2)$$

$$3ax_1^{\ 2} + 2bx_1 = f(x_2) = p_2'(x_2)$$

These equations can be solved for a and b.

Now $f_1(x) = 3ax^2 + 2bx = 0$ when $x = 0$ or when $x = \frac{-2b}{3a}$

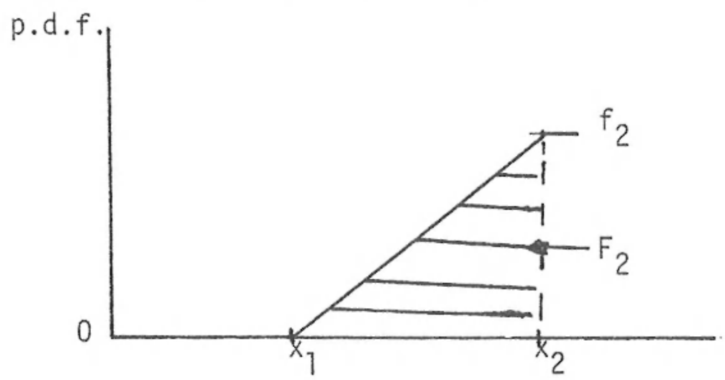
If $0 < \frac{2b}{3a} < x_2$ $f_1(x)$ is negative in $(0,x_2)$ and a further adjustment is needed. In this case, we assume a linear form for the density function from point x_1 in $(0, x_2)$ i.e.

$$p_1'(x) = 2ex + f \quad x > x_1$$
$$= 0 \text{ if } 0 < x \leq x_1$$

Then $p_1(x) = 0$ if $0 < x \leq x_1$

$$= ex^2 + fx + g \text{ if } x_1 < x < x_2$$

The situation can be depicted graphically as follows



So $p_1'(x) = \dfrac{f_2}{x_2 - x_1} x - \dfrac{f_2 x_1}{x_2 - x_1}$ and $F_2 = f_2 \dfrac{(x_2 - x_1)}{2}$

$$x_1 = x_2 - \frac{2F_2}{f_2} > 0 \text{ if } x_2 > \frac{2F_2}{f_2}$$

Since $p_1(x_1) = 0$

$$e = \frac{f_2}{2(x_2 - x_1)}$$

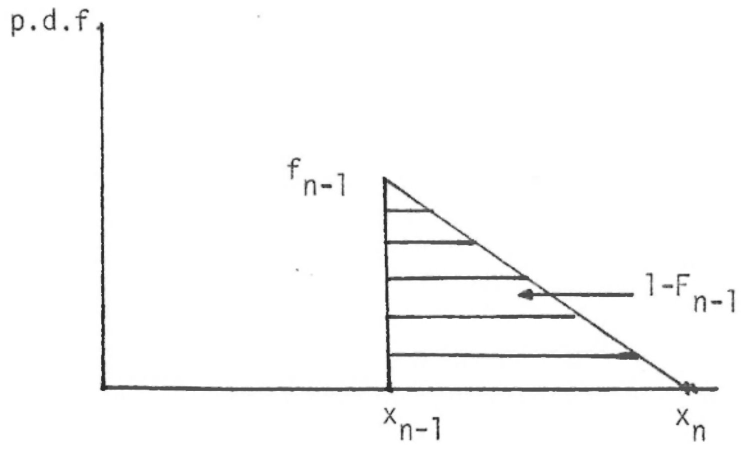
$$f = \frac{f_2 x_1}{x_2 - x_1}$$

$$g = ex_1^2 - fx_1 = -\frac{f_2 x_1^2}{2(x_2 - x_1)} + \frac{f_2 x_1^2}{(x_2 - x_1)} = \frac{f_2 x_1^2}{2(x_2 - x_1)}$$

In most cases, the left tail will be satisfactorily fitted by making the first or the second adjustment. However, neither the condition for the first adjustment $-\frac{2b}{3a}$ not in $(0, x_2)$ nor that for the second $x_2 > \frac{2F_2}{f_2}$ may be satisfied. In that case, we restart the whole cubic spline function calculation with $x_1 = \frac{x_2}{2}$ and try again. If that fails, try with $x_1 = \frac{3x_2}{4}$ etc., halving the distance between x_1 and x_2 until the tail is satisfactorily fitted. Then $p_1'(x_1) = 0$ and $p'(x)$ is continuous at $x = x_2$ with $p_1(x_2) = F(x_2)$.

(b) Last class :

We assume that the density function has a linear form in the open class which extends from x_{n-1} to x_n, as shown below



$$1 - F_{n-1} = \frac{(x_n - x_{n-1})}{2} \; f_{n-1}$$

$$x_n = x_{n-1} + 2\left(\frac{1-F_{n-1}}{f_{n-1}}\right) > x_{n-1} \quad \text{if } f_{n-1} > 0$$

$$p'_{n-1}(x) = \frac{x_n}{x_n - x_{n-1}} \; f_{n-1} - \frac{x}{x_n - x_{n-1}} \; f_{n-1}$$

$$= 2\,px + q$$

$$p = \frac{f_{n-1}}{2(x_n - x_{n-1})}$$

$$q = \frac{f_{n-1}\, x_n}{x_n - x_{n-1}}$$

$$P_{n-1}(x) = px^2 + qx + r$$

$$p_{n-1}(x_n) = 1$$

$$- \frac{f_{n-1}\, x_n^{\,2}}{2(x_n - x_{n-1})} + \frac{f_{n-1}\, x_n^{\,2}}{(x_n - x_{n-1})} + r = 1$$

$$r = 1 - \frac{f_{n-1}\, x_n^{\,2}}{2(x_n - x_{n-1})}$$

Appendix II.

COMPUTATION OF GENERAL INEQUALITY MEASURES i_1, i_2, i_3, i_4 AND i_6.

(i) $$1-i_1 = \frac{\bar{x}^2}{s_x^2+\bar{x}^2}$$

so $$i_1 = 1 - \frac{\bar{x}^2}{s_x^2+\bar{x}^2} = \frac{s_x^2}{s_x^2+\bar{x}^2} = \frac{1}{1+c^2},$$

where $\bar{x} = \int x f(x)dx$, the arithmetic mean

$s_x^2 = \int (x-\bar{x})^2 f(x)dx$, the variance of incomes

and $c = \frac{s_x}{\bar{x}}$, the coefficient of variation of incomes.

(ii) $$1-i_2 = \frac{1}{s_z^2+1}$$

so $$i_2 = \frac{s_z^2}{s_z^2+1}$$

where s_z is the standard deviation of income-power z (=ln x)

Now $$s_z^2 = \int \left[\ln(\frac{x}{\bar{x}_g})^2\right] f(x)dx$$

where $\bar{x}_g = \exp\{\int \ln x\, f(x)dx\}$

(iii) $$1 - i_3 = \frac{\bar{x}_g}{\bar{x}} \quad \text{so} \quad i_3 = 1 - \frac{\bar{x}_g}{\bar{x}}$$

(iv) $$i_4 = 1 - \frac{1}{\bar{x}}\left[\int x^{1-\varepsilon} f(x)dx\right]^{\frac{1}{1-\varepsilon}}$$

Now $\varepsilon = 2$ so $$i_4 = 1 - \frac{1}{\bar{x}}\left[\int x^{-1} f(x)dx\right]^{-1} = 1 - \frac{1}{\bar{x}\int \frac{1}{x} f(x)dx}$$

(v) Denote the Theil measure by T and the N incomes by $x_1, \ldots x_n$.

Then $T = \sum_{r=1}^{n} \frac{x_r}{N\bar{x}} \ln \left(\frac{x_r}{\bar{x}}\right)$

This varies from 0 to ln N, so define I_6 as

$$I_6 = \frac{T}{\ln N} = \sum_{r=1}^{n} \frac{\frac{x_r}{\bar{x}} \ln \left(\frac{x_r}{\bar{x}}\right)}{N \ln N}$$

Now, following Champernowne,

let $i_6 = 1 - N^{-I_6}$

$$\ln (1-i_6) = -I_6 \ln N$$

$$= \frac{-1}{N} \sum_{r=1}^{n} \frac{x_r}{\bar{x}} \ln \left(\frac{x_r}{\bar{x}}\right)$$

$$= \int_o^\infty \left(\frac{x}{\bar{x}}\right) \ln \left(\frac{x}{\bar{x}}\right) f(x)dx \quad \text{in the continuous case}$$

$$= \ln \bar{x} - \frac{\int_o^\infty x \ln x \ f(x)dx}{\bar{x}}$$

$$1 - i_6 = \frac{\bar{x}}{\exp \left\{\int_o^\infty \frac{x \ln x}{\bar{x}} \ f(x)dx\right\}}$$

$$i_6 = 1 - \frac{\bar{x}}{\exp\left\{\int_o^\infty \frac{x \ln x}{\bar{x}} \ f(x)dx\right\}}$$

Appendix III

PROOF OF FORMULAE FOR I, G AND P - POVERTY ANALYSIS.

$$F(x) = Ax^{\gamma'} \quad 0 \leq x \leq y$$

$$F(y) = 1 = Ay^{\gamma'}$$

$$A = \frac{1}{y^{\gamma'}}$$

$$f(x) = \gamma' A x^{\gamma'-1} = \frac{\gamma' x^{\gamma'-1}}{y^{\gamma'}} \quad 0 \leq x \leq y$$

$$= 0 \text{ elsewhere}$$

The mean income $\bar{x}_y$ among the households below the MLL is

$$\bar{x}_y = \int_0^y \frac{x\,\gamma' x^{\gamma'-1}}{y^{\gamma'}} = \frac{\gamma'}{y^{\gamma'}} \int_0^y x^{\gamma'}\,dx = \frac{\gamma'}{y^{\gamma'}} \frac{y^{\gamma'+1}}{\gamma'+1}$$

$$\text{Hence, } I = \frac{y-\bar{x}_y}{y} = \frac{y - \frac{\gamma' y}{\gamma'+1}}{y} = \frac{\gamma'+1-\gamma'}{\gamma'+1} = \frac{1}{\gamma'+1}$$

$$\text{Let} \quad \phi(t) = \frac{1}{\bar{x}_y} \int_0^t x f(x)dx = \frac{\gamma'+1}{\gamma' y} \int_0^t \frac{\gamma' x^{\gamma'}}{y^{\gamma'}}\,dx$$

$$= \frac{\gamma'+1}{\gamma' y} \frac{\gamma'}{y^{\gamma'}} \frac{t^{\gamma'+1}}{\gamma'+1} = \frac{t^{\gamma'+1}}{y^{\gamma'+1}}$$

$$G = 1 - 2\int_0^y \phi(t)f(t)dt$$

$$= 1 - 2\int_0^y \frac{t^{\gamma'+1}}{y^{\gamma'+1}} \frac{\gamma' t^{\gamma'-1}}{y^{\gamma'}}$$

$$= 1 - \frac{2\gamma'}{y^{2\gamma'+1}} \int_0^y t^{2\gamma'}\,dt$$

$$= 1 - \frac{2\gamma'}{y^{2\gamma'+1}} \frac{y^{2\gamma'+1}}{2\gamma'+1}$$

$$= 1 - \frac{2\gamma'}{2\gamma'+1} = \frac{1}{2\gamma'+1}$$

$$P = H\left[I + (1-I)G\right] = H\left[\frac{1}{\gamma'+1} + \frac{\gamma'}{\gamma'+1} \frac{1}{2\gamma'+1}\right]$$

$$= H \frac{2\gamma'+1+\gamma'}{(\gamma'+1)(2\gamma'+1)} = H \frac{3\gamma'+1}{(\gamma'+1)(2\gamma'+1)}$$

NOTES TO APPENDICES.

1 Carnahan and Wilkes *op. cit.*, p. 309.

2 Champernowne, *op. cit.*, p. 791.

3 *Ibid.*

4 *Ibid.*

5 *Ibid.*

6 H. Theil, *op. cit.*, p. 91.

7 D.G. Champernowne: *Notes on the formulae quoted in 'A comparison of measures of inequality of income distribution'* (mimeo) equation 1.5.13

DEVELOPMENT STUDIES RESEARCH GROUP

Publications, Papers and Proceedings

BOOKS

1. CHARLES SIMKINS AND DUNCAN CLARKE, *Structural Unemployment in Southern Africa*, Natal University Press, 1978. R3,75 (Order from the Press at P.O. Box 375, Pietermaritzburg, 3200).

2. CHARLES SIMKINS AND COSMAS DESMOND, (eds.), *South African Unemployment:A Black Picture*, Development Studies Research Group and Agency for Industrial Mission, 1978. Price: R3.

WORKING PAPERS

			COST
1.	D.G. CLARKE	Foreign African Labour Supply in South Africa, 1960-1977. 62p. April 1977	2,00
2.	D.G. CLARKE	The South African Chamber of Mines: Policy and Strategy with Reference to Foreign African Labour Supply. 45p. April 1977	1,00
3.	C.E.W. SIMKINS	Conditions of Work in Contemporary South African Industry: A Study of Minimum Standards. 26p. April 1977	1,00
4.	D.G. CLARKE	State Policy on Foreign African Labour in South Africa: Statutory, Administrative and Contractual Forms. 27p. September 1977	1,00
5.	COSMAS DESMOND	Limehill Revisited: a Case Study of the Longer-term Effects of African Resettlement. 28p.Jan'78	1,00
6.	C.E.W. SIMKINS	Income Distribution among Settled Urban African Households in South Africa:1970 & '75.59p.Mar.78.	2,00
7.	C.E.W. SIMKINS & D.C.HINDSON	The Division of Labour in South Africa, 1969-1977 Feb. 1979	2,00
8.	C.E.W SIMKINS	South African Development in International Perspective 1950-1975 44p. Sept. 1979	2,00
9.	C.E.W. SIMKINS	The Distribution of Personal Income among income recipients in South Africa, 1970 & 1976.......	1,00
10.	C.E.W. SIMKINS	Agricultural Production in the African Reserves of South Africa, 1918-1969..28 p...Dec. 1979..	1,00

DISCUSSION PAPERS

1.	D.G. CLARKE	Inflation and Subsistence Wages: Revised Estimates of the Urban Poverty Datum Line in Rhodesia for Sept. 1976. 10p	0,50
2.	D.G. CLARKE	Poverty and Unemployment in Nqutu. 9p..........	0,50
3.	C. SIMKINS	Development economics and radical theories of the peripheral state .. 17p	0,50
4.	C. SIMKINS	Male African Unemployment in the Richards Bay area	0,50
5.	C. SIMKINS	The Current Population Survey and South African Unemployment: Some Puzzles	0,50
6.	M. MCGRATH	A Comment on the Current Population Survey, Income Distribution and Poverty	0,50
7.	C. SIMKINS	On Quantitative methods in the social sciences .	0,50

PROCEEDINGS

WORKSHOP ON UNEMPLOYMENT AND LABOUR REALLOCATION IN SOUTHERN AFRICA, DSRG, PIETERMARITZBURG, 18/19 Mar. 1977 (8 papers) 5,50

Copies of all items listed may be purchased. The DSRG is also interested in exchanging literature with researchers in related fields. Proposals in this connection, orders for papers or request for further information should be addressed to Sec. Development Studi Research Group. Dept. Economics, University of Natal, P.O. Box 375, Pietermaritzburg.32